

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO



Uma Visão Sobre uma Forma de Retenção de Clientes nos Seguros Automóveis

Beatriz Ferraz Marreiros

Mestrado de Matemática Aplicada à Economia e Gestão

Trabalho de Projeto orientado por:
Professor Doutor João José Ferreira Gomes

2017

“Se não receio o erro, é porque estou sempre disposto a corrigi-lo”
(Bento de Jesus Caraça, Matemático Calipolense, 1901-1948)

AGRADECIMENTOS

Agradeço ao meu orientador Professor João Gomes e à Professora Teresa Alpuim por todo o apoio prestado assim como por todas as sugestões que levaram ao sucesso da realização desta tese.

Agradeço também aos meus colegas de trabalho que me apoiaram e que, com a sua experiência, me ajudaram a enriquece-la. Em especial, à Joana Camilo pela sua disponibilidade. Um agradecimento especial também ao meu atual chefe Rui Faria. Além de me aconselhar e de me disponibilizar muito tempo, também se mostrou sempre acessível a qualquer questão sobre o tema que eventualmente surtisse.

Porque para o sucesso da realização desta tese não é só necessário existir conhecimento teórico e prático agradeço à minha mãe, Fátima Ferraz, e à minha irmã, Carolina Ferraz Marreiros, pelo apoio incondicional que me tornou mais forte e pela contínua motivação (mesmo nos momentos mais complicados). Agradeço também ao meu pai, Paulo Marreiros, que me proporcionou a oportunidade de fazer este mestrado.

Ao Instituto de Odivelas que me deu todas as bases e valores para atingir este objetivo. Em especial às amizades que lá criei que se lembraram sempre de me fazer esquecer a tese.

RESUMO

São vários os fatores que contribuem para o crescimento de uma companhia de seguros, nomeadamente, os critérios de avaliação interno para um bom saneamento de carteira e uma forte base de fidelização dos seus clientes.

No crescimento de uma companhia de seguros é tão importante existir um bom saneamento como também uma boa fidelização dos clientes. Uma vez que são estes clientes que vão dar à companhia de seguros estabilidade.

O sucesso dos seguros baseia-se no conceito de mutualização de risco, isto é, redistribuição do risco por um conjunto de pessoas, neste caso, pelo conjunto da carteira de apólices em vigor. Logo para a sobrevivência de uma companhia de seguros é essencial eliminar o risco (saneamento) e aumentar a carteira de apólices por onde se redistribuir o mesmo. Diminuindo assim a responsabilidade que recai diretamente sobre a companhia de seguros.

Numa primeira fase é feita uma avaliação à carteira relativamente ao tempo de apólice segura. Esta avaliação foi feita em três níveis: na carteira em geral, nas apólices de colaboradores e nas apólices de não colaboradores. Com estes três níveis analisou-se a evolução da carteira em termos de crescimento da mesma, de taxa de anulação e de número de apólices criadas anualmente, onde se detetou uma fraca retenção de clientes.

De seguida procurou-se acompanhar a proporção de anulação das apólices a partir do ano de criação das mesmas. Através desta análise tem-se uma noção da evolução das apólices anuladas. Mais uma vez se prova que a não retenção de clientes é um problema.

Numa segunda fase desenvolve-se um modelo que quantifica a probabilidade de uma apólice anular através do modelo de regressão logística.

O que se pretende é resolver os problemas identificados inicialmente. Através da aplicação de descontos para as apólices com maior probabilidade de anular e agravamentos em caso contrário.

Por fim é feita uma análise descritiva em termos de sinistralidade, características da apólice e do tomador de seguro para se “conhecer” melhor a carteira que está a ser estudada. Terminando com uma aplicação prática dos resultados anteriores.

PALAVRAS-CHAVE:

Seguro Automóvel, probabilidade, modelo de regressão logística, sinistralidade, retenção

ABSTRACT

In the growth of an insurance company as important as a good selection of insurance policies is as well a good client loyalty. Once this clients will give stability to the company.

The success of the insurance concept it's based on the redistribution of the risk in a set of people, in this case, all of the insurance policies in force. So for the survival of insurance company such important as eliminate the risk (separation between bad/good insurance policies), is to improve the number of insurance policies to redistribute the same.

Reducing the liability that falls directly on the insurance company.

In a first stage was made a review of the insurance policies duration.

This review was made in three levels: in the portfolio in general, on the employee portfolio and on the no-employee portfolio. With this three levels it was analyzed the evolution of the portfolio in terms of growth, cancellation rate, and numbers of insurance policies created annually. Where was sensed a weak client retention.

Subsequently we tried to follow the ratio of canceled insurance policies based on the year they were created. Through this analysis we have a notion of the evolution of the insurance policies canceled. Once again it was proved that client retention is a problem.

In a second stage we developed a model that quantifies the probability of annulation of insurance policies through the logistic regression model.

The aim is to resolve the problems identified initially, through the application of discounts for the insurance policies most likely to abort, and charge more on the opposite case.

For last we made a descriptive analysis in terms of accident rates, features of the insurance policy and policyholder, to know better the portfolio that is being studied.

Ending with a practical application of the previous results.

KEYWORDS:

Car insurance, probability, logistic regression model, accidents, retention

ÍNDICE

1. FUNCIONAMENTO GERAL DOS SEGUROS.....	19
Seguro Automóvel.....	19
Regularização de Sinistros	22
2. REGRESSÃO LOGÍSTICA	25
Regressão Logística Simples.....	25
Regressão Logística Múltipla.....	26
Escolha do Modelo.....	27
Teste de razão de Verossimilhança.....	27
Teste Wald.....	27
Seleção de Variáveis	28
Qualidade do Modelo Final.....	29
Resíduos de Pearson.....	29
Resíduos de Deviance	30
Teste de Hosmer-Lemeshow	30
Curva ROC.....	31
3. ANÁLISE GERAL À CARTEIRA.....	33
Base de Dados	33
Apólices (ap)	33
Sinistros (ps).....	35
Variáveis Extra	36
Análise Geral à Carteira	36
Fidelização	37
Constituição da Carteira	37
Apólices de Colaborador Vs Não Colaborador	38
Taxa de Anulação.....	41
Apólices Criadas.....	43
Crescimento Carteira de Apólices	45
Proporção de Apólices Anuladas	47
Modelos de Regressão Logística - Estimar a Probabilidade de Anulação de ap.....	50
Análise Simples.....	51
Análise Múltipla.....	53
Qualidade do modelo final	55
Restantes Modelos Finais.....	56

Sinistralidade	58
Aplicações	63
4. CONCLUSÃO	65
5. BREVE GLOSSÁRIO DE SEGUROS	67
6. BIBLIOGRAFIA	73
7. ANEXOS	75
Anexo 1 – Tabela de Desvalorização da Viatura Segura	75
Anexo 2 – Tabela de Saneamento considerada	75
Anexo 3 – Franquias por coberturas.....	76
Anexo 4 – Coberturas e Modalidades de ap	76
Anexo 5 – Comparação de modelos (P1, P2 e P3)	78

ÍNDICE DE FIGURAS

Figura 1.1 Distribuição do prejuízo com e sem franquia	21
Figura 2.1 Três casos tipo da curva ROC.....	32
Figura 3.1 Tempo de ap segura (em geral).....	38
Figura 3.2 Tempo de ap segura entre ap de colaborador e não colaborador	39
Figura 3.3 Percentagem de ap (geral, colaborador e não colaborador) por tempo seguro	40
Figura 3.4 Percentagem de ap (geral, colaborador e não colaborador) por intervalo de tempo	40
Figura 3.5 Variação da taxa anual entre anos consecutivos I.....	42
Figura 3.6 Taxa de anulação anual	42
Figura 3.7 Variação da taxa anual entre anos consecutivos II	43
Figura 3.8 Quantidade de ap criadas anualmente.....	44
Figura 3.9 Variação do número de apólices criadas anualmente	44
Figura 3.10 Crescimento da carteira de ap	45
Figura 3.11 Variação do crescimento da carteira de ap	46
Figura 3.12 Comparação entre classes da taxa de anulação, criação de apólices e crescimento da carteira.....	46
Figura 3.13 Proporção de ap anuladas.....	47
Figura 3.14 Variação da proporção de ap anuladas.....	48
Figura 3.15 Fase (1) e fase (2) na anulação de ap	48
Figura 3.16 Percentagem das ap em vigor em cada uma das BD.....	51
Figura 3.17 Proporção de ap anuladas por número de sinistro conhecido	52
Figura 3.18 Curva ROC para P1	55
Figura 3.19 Comparação dos valores observados e estimados - P1	55
Figura 3.20 Curva ROC para P2	56
Figura 3.21 Comparação dos valores observados e estimados - P2	56
Figura 3.22 Curva ROC para P3	57
Figura 3.23 Comparação dos valores observados e estimados - P3	57
Figura 3.24 Percentagem de tipo de ps participados/reclamados à seguradora.....	60
Figura 3.25 Percentagem de opção de ap por tipo de ps	61
Figura 3.26 Percentagem de opção de ap por tipo de ps	61
Figura 3.27 Ocorrência de sinistros mensal	62

ÍNDICE DE TABELAS

Tabela 3.1 Percentagem de ap (geral, colaborador e não colaborador) por tempo seguro	40
Tabela 3.2 Distribuição de apólice segura por intervalo de tempo (geral, colaborador e não colaborador)	40
Tabela 3.3 Proporção de apólices anuladas por ano de criação - acumulada	47
Tabela 3.4 Proporção de ap anuladas - não acumulada	49
Tabela 3.5 AIC de cada processo utilizado para seleção de variáveis.....	53
Tabela 3.6 Ajustamento de mod2 – P1	54
Tabela 3.7 Coberturas mais cara (custo médio) - sinistralidade.....	58
Tabela 3.8 Custo médio de cada risco por cobertura - sinistralidade	59
Tabela 3.9 Custo médio e percentagem por ps	60
Tabela 3.10 FN e FP nos modelos P2 e P3. Considerando um ponto de corte de 70% para descontos e 2% para agravamentos.....	63
Tabela 7.1 Tabela de desvalorização da viatura segura.....	75
Tabela 7.2 Regras de saneamento	75
Tabela 7.3 Franquia aplicada a cada cobertura contratada.....	76
Tabela 7.4 Opções de ap disponível	77
Tabela 7.5 Comparação dos três modelos: P1, P2 e P3.....	78

LISTA DE SIGLAS E ABREVIATURAS

AP	Apólice
ASF	Autoridade de Seguros de Portugal
APS	Associação Portuguesa de Seguradores
AV	Atos de Vandalismo
BD	Base de Dados
BM	Bónus-Malus
CCC	Choque, Capotamento ou Colisão
DP	Danos Próprios
FGA	Fundo de Garantia Automóvel
FR	Furto ou Roubo
IDS	Indemnização Directa ao Segurado
INEM	Instituto Nacional de Emergência Médica
IRE	Incêndio, Raio ou expulsão
PS	Processo Sinistro
PU	Privação de Uso
QIV	Quebra Isolada de Vidro
RCO	Responsabilidade Civil Obrigatória
SNBPC	Serviço Nacional de Bombeiros e Protecção Civil
TPR	Tabela Prática de Responsabilidades
VS	Viatura de substituição

1. FUNCIONAMENTO GERAL DOS SEGUROS

Segundo Moitinho de Almeida o contrato de seguro define-se “como aquele em que uma das partes, o segurador, compensando segundo as leis da estatística um conjunto de riscos por ele assumidos, se obriga mediante o pagamento de uma soma determinada, no caso de realização de risco, a indemnizar. (...)” (Almeida, 1971).

O seguro surgiu pela necessidade de transferência de risco, funcionando como um método de proteção à incerteza. O segurado através da troca de um montante monetário (o prémio) transfere para a seguradora a responsabilidade do pagamento consequente de um sinistro. Deste modo fica seguro de eventuais perdas monetárias de elevado montante.

O principal conceito para que o seguro seja benéfico para ambas as partes baseia-se na redistribuição do risco. Deste modo chegamos ao principal conceito nos seguros: Mutualidade.

A mutualidade é o processo de distribuir por muitos os prejuízos sofridos por poucos. Um custo que podia ser devastador para apenas um indivíduo reduz significativamente se divididos por muitos.

“Os seguros protegem pessoas e empresas contra o risco de eventos imprevisíveis. São um mecanismo de transferência de risco baseado na mutualização, ou partilha, de risco. As perdas de poucos são pagas pelos prémios de muitos, calculados a partir do risco de cada indivíduo ou entidade.”
(Insurance Europe, 2010)

Devido à desproporção entre o eventual dano consequente de um sinistro e a capacidade indemnizatória do responsável pelo mesmo, tornou-se obrigatório a necessidade de contratar o seguro automóvel. Assim redistribui-se o valor dos prémios de seguro de todas as apólices pelas indemnizações dos danos. Estamos perante a função social do contrato de seguro, à qual está subjacente os princípios de interesse público.

O risco é essencial no contrato de seguro e resulta da possibilidade da ocorrência de um sinistro que origine um dano patrimonial não evitável. O risco reflecte então o interesse protegido (estabelecendo uma relação económica entre uma pessoa e uma coisa exposta ao risco em causa).

“Risco é a probabilidade de perda”
(Real, 2009)

Seguro Automóvel

Para celebrar uma apólice de seguro é necessário passar primeiro pelo processo de seleção, ou seja, a subscrição. Esta fase envolve uma tomada de decisão seletiva de riscos a aceitar e qual o valor de prémio necessário cobrar, tendo em conta o perfil e as coberturas pedidas. O principal objetivo da subscrição é então garantir que a carteira de apólice se mantém além de lucrativa crescente.

Note-se que os critérios de seleção se baseiam em análises estatísticas da gestão de carteira de modo a traçar um perfil para a ocorrência do risco segurado. Aqui, calcula-se qual o valor do prémio a cobrar.

Apesar de o mesmo acontecimento poder levar a perfis muito idênticos o processo de subscrição varia de seguradora para seguradora. Esta avaliação depende sempre da sua posição relativamente ao risco, ou seja, do nível de risco a que a seguradora está disposta a aceitar.

“O processo de subscrição beneficia os tomadores de seguro. Quanto mais informação existir sobre um risco individual, mais o prémio pode ser adaptado a esse risco”
(Insurance Europe, 2010)

É na subscrição que além de se celebrarem contratos de seguro se actualizam (gestão de risco) os valores de prémios cobrados (bonús/malus) e capital seguro (desvalorização do valor da viatura segura).

Suponhamos que o capital seguro é calculado de acordo com o valor da viatura a segurar:

- Veículo novo: o valor do seguro corresponderá ao preço de venda, em Portugal, no mês e ano da sua matrícula. Considerando todos os impostos e encargos aplicáveis, sem descontos comerciais e acrescido de os equipamentos que não integrem o modelo do veículo adquirido
- Veículo usado: o valor seguro corresponderá ao valor de aquisição inicial atualizado de acordo com a tabela de desvalorização – anexo 1

O sistema de *bonús/malus* é um sistema que regula o valor do prémio do seguro. Segundo o qual, de acordo com anexo 2, o prémio a pagar no período seguinte será atualizado.

Em resumo o sistema de *BM* corresponde a um fator de tarificação à *posteriori*. Onde se bonifica o segurado que não declarou sinistros e se penaliza o inverso.

Conforme dito anteriormente é necessário traçar um perfil e ajustar uma tarificação adequada ao mesmo. Torna-se fundamental que exista na companhia de seguros uma base de dados fiável e com um histórico mínimo de 5 anos.

Para celebrar um contrato de apólice de seguro é necessário o pagamento de um montante monetário a que se dá o nome de prémio. Parte deste valor é calculado pela igualdade entre quanto o segurado está disposto a pagar para delegar o risco e quanto a seguradora está disponível a cobrar para assumir esse mesmo risco, ou seja:

Seja $u(x)$ o valor que a pessoa atribui ao facto de ganhar o montante x (Teoria da Utilidade).

Perspetiva do segurado: a igualdade entre ele próprio assumir o risco e delegar o mesmo à seguradora em troca de pagar o montante P_1 é dado como:

$$E[u_1(w_1 - x)] = u_1(w_1 - P_1)$$

$u_1(x)$: função utilidade para o segurado

w_1 : riqueza do segurado

P_1 : prémio máximo para o segurado delegar o risco aleatório de x

x : prejuízo consequente do acontecimento seguro (considera-se como sendo uma variável aleatória uma vez que se trata de um acontecimento inesperado)

Perspetiva da seguradora: igualdade entre não assumir o risco e assumir o mesmo em troca de um montante P_2 a receber pelo segurado é dado como:

$$u_2(w_2) = E[u_2(w_2 + P_2 - x)]$$

$u_2(x)$: função utilidade para a seguradora

w_2 : riqueza da seguradora

P_2 : prémio mínimo para a seguradora assumir o risco aleatório de x

x : prejuízo consequente do acontecimento seguro (considera-se como sendo uma variável aleatória uma vez que se trata de um acontecimento inesperado)

Assim a apólice só é praticável caso: $P_1 \geq P_2 \geq \mu$ onde $\mu = E(x)$ representa o prémio puro, ou seja, o prejuízo esperado.

O ideal seria que o prêmio fosse calculado apenas segundo a lógica acima explicada. No entanto existem taxas que têm de ser pagas a várias entidades como:

- FGA: trata-se de uma entidade que se destina a regularizar sinistros dentro de circunstâncias muito específicas;
- INEM;
- Imposto de Selo;
- Prevenção Rodoviária ou Serviço Nacional de Bombeiros e Proteção Civil;

No sentido de a seguradora diminuir os seus custos surge o conceito de franquia. A franquia corresponde ao valor de reparação consequente de um sinistro que fica a cargo do tomador do seguro. Esta pode ter varias formas de funcionamento: ser obrigatória ou facultativa, ter um valor fixo ou até mesmo uma percentagem do valor do capital seguro ou do dano.

A franquia também desempenha um papel preventivo uma vez que o tomador de seguro tenderá a ser mais cuidadoso, podendo desta forma evitar a ocorrência de sinistros. Esta é a forma de responsabilizar o cliente por uma parte do prejuízo causado.

Este método (a aplicação de franquia) é uma forma de gestão de risco aplicada pela companhia de seguros.

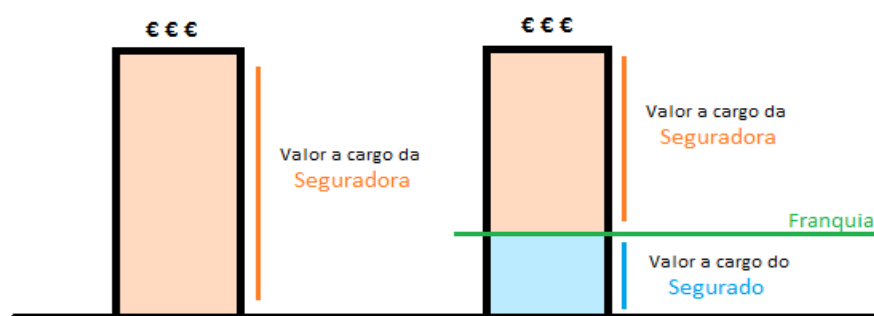


Figura 1.1 Distribuição do prejuízo com e sem franquia

A franquia traz vantagens não só para a seguradora que vê os seus custos reduzidos mas também para os segurados. Uma vez que estes ficam responsáveis por parte dos prejuízos podendo até usufruir de um prémio do seguro mais baixo por existir franquia.

Em geral, a franquia apenas se aplica a coberturas onde o condutor poderá ter culpa direta no sinistro – anexo 3.

A atividade seguradora garante parte da segurança económica dos seus clientes. Para o seguro automóvel existem três modalidades: Seguro de Responsabilidade Civil Automóvel, Seguro de Danos Próprios ou Seguro de Danos Próprios sem Choque (que no senso comum são conhecidos como mínimos, máximos e médios respetivamente) – ver anexo 4:

- Seguro de Responsabilidade Civil Automóvel é a única cobertura obrigatória e comum a todas as apólices automóvel, com um capital mínimo obrigatório. Esta cobertura garante todos os danos causados a terceiros;
- Seguro de Danos Próprios sem choque: além da cobertura de Responsabilidade Civil (que é obrigatória) cobre também casos de furto ou roubo, atos de vandalismo, incêndio, raio ou explosão, quebra isolada de vidro (QIV) e fenómenos da Natureza;
- Seguro de Danos Próprios é conhecido como o seguro contra todos os riscos uma vez que cobre qualquer sinistro mesmo que o responsável seja o próprio condutor da via-tura;

O resumo da distribuição das coberturas disponíveis pelas 3 modalidades de apólices existentes e as várias possibilidades de capitais seguros considerados encontra-se no anexo 4.

Considera-se as três opções (máximos, médios e mínimos) como:

- Opção mínimos: Responsabilidade Civil (“contra terceiros”), pelo mínimo legalmente exigido, ainda que possa ser reforçada com outras coberturas opcionais;
- Opção médios: reforça a opção anterior (mínimos) e em uma componente de danos próprios;
- Opção máximos: opção com o maior nível de proteção uma vez que inclui reforço total da componente de danos próprios;

Apesar de ser obrigatório nem todas as viaturas se encontram seguras. Nestas situações o Fundo de Garantia Automóvel (FGA) funciona como um mecanismo que garante o pagamento das indemnizações.

Assim toda a reparação dos danos causados por um responsável desconhecido ou sem seguro válido à data do sinistro é garantida pelo Fundo de Garantia Automóvel.

Note-se que o FGA assume as indemnizações mas exige posteriormente, aos responsáveis, o reembolso do sinistro.

Regularização de Sinistros

As companhias de seguros têm duas maneiras de tomar conhecimento da ocorrência de um sinistro:

- Reclamação: se participado pelo condutor da viatura terceira;
- Participação: se participado pelo condutor da viatura segura;

Deste modo para participação/reclamação de sinistro o ideal é o correto preenchimento da Declaração Amigável de Acidente Automóvel (D.A.A.A.) – notificação. Onde se dá a conhecer os factos e a identificação dos intervenientes numa única visão.

No entanto nem sempre existe disponível uma DAAA pelo que também se considera participação/reclamação de sinistro qualquer documento onde esteja identificado: a data e o local do sinistro, as circunstâncias do mesmo e com pelo menos uma assinatura de um dos condutores.

Para a regularização de sinistros existem várias convenções que visam a facilitar. Como é o caso da convenção IDS (à qual a maioria das seguradoras aderiu). Deste modo após a ocorrência de sinistro este pode ser regularizado através de processo tradicional (não IDS) ou através de qualquer outro acordo como o caso da convenção IDS.

Considere-se apenas a convenção IDS para a regularização de sinistros além do processo tradicional.

Quanto à regularização de sinistros, o principal objetivo é possibilitar que o segurado receba da sua seguradora a indemnização a que tem direito o mais rápido possível.

De um modo geral a regularização de sinistros (para danos materiais) funciona do seguinte modo:

1. contacto com o tomador de seguro ou terceiro lesado para marcação de peritagem
 - a. Caso necessário fazer as averiguações precisas do sinistro, auto policial...;
2. Comunicação da definição de responsabilidade ao tomador de seguro e terceiro lesado (assunção ou não de responsabilidade do sinistro);
3. Se a responsabilidade for assumida prossegue-se à reparação da viatura;
4. Caso seja declinada a responsabilidade o processo termina no ponto 2;

A convenção IDS é um serviço prestado no âmbito do seguro obrigatório de responsabilidade civil automóvel. De forma a facilitar a rápida regularização de sinistros automóvel ocorridos em Portugal.

Esta convenção pode permitir resolver rapidamente cerca de 80% dos sinistros de responsabilidade civil.

As condições necessárias para um sinistro poder ser regularizado via IDS são:

- Ocorrência do sinistro em território português;
- Envolvidas apenas duas viaturas;
- Embate direto entre os dois veículos;
- As duas viaturas tenham seguro obrigatório de responsabilidade civil válido à data do sinistro;
- Declaração Amigável de Acidente Automóvel (DAAA) devidamente preenchida;
- As duas viaturas envolvidas têm de ter matrícula portuguesa;
- As duas apólices de seguros têm de ter sido emitidas por seguradoras sediadas em Portugal e aderentes à convenção;
- Danos inferiores a 15.000 € por viatura;
- Não existência de danos corporais mesmo que ligeiros;

Caso o sinistro não possa ser enquadrado na convenção IDS então o mesmo é regularizado por processo tradicional - seguindo o esquema anteriormente explicado.

Esta convenção enquadra as circunstâncias indicadas na DAAA na Tabela Prática de Responsabilidades (TPR) e não no código da estrada. A TPR é um conjunto de gráficos ilustrativos de casos tipo de sinistro, através dos quais já estão pré-definidas as percentagens de responsabilidade para cada um dos intervenientes.

Resumidamente, a regularização dos sinistros via IDS é igual ao esquema inicialmente explicado. As únicas diferenças são:

- É dado um acordo entre companhias de seguros sobre a quem fica a responsabilidade antes da regularização do sinistro em si – o que torna todo o processo mais rápido;
- É aplicado a TPR e não o código da estrada;

2. REGRESSÃO LOGÍSTICA

Regressão Logística Simples

No modelo de regressão linear assume-se que:

$$2.1 \quad E[Y|x] = \beta_0 + \beta_1 x + \varepsilon$$

Onde β_0 e β_1 são parâmetros a estimar e ε 's são variáveis aleatórias que representam o erro (ou seja, devolvem o desvio da observação em relação à média condicional) com a média nula e a variância constante, não correlacionadas, isto é, dada uma amostra (x_i, y_i) $i = 1, \dots, n$ da População:

- Condições de *Gauss-Markov*
 - $E(\varepsilon_i) = 0$
 - $Var(\varepsilon_i) = \sigma^2$
 - $E(\varepsilon_i \varepsilon_j) = 0, se i \neq j$
- Pode ainda ser exigido que ε_i 's sejam independentes e com uma distribuição Normal, $N(0, \sigma^2)$

Na regressão linear a observação da variável resposta é dada como:

$$2.2 \quad y = E[Y|x] + \varepsilon$$

Uma vez que a variável resposta é binomial (toma apenas o valor 0 ou 1) terá de ser aplicado regressão binária (no nosso caso utilizaremos a regressão logística) sendo esta a maior diferença entre a regressão linear e a regressão binária. Deste modo as condições anteriores não se aplicam quando a variável é binária, devendo-se a mesma representar como:

$$2.3 \quad y = \pi(x) + \varepsilon$$

Considerando $\pi(x) = E[Y|x]$

No entanto, uma vez que a variável resposta tem distribuição *Bernoulli* ($1, P(x)$), com probabilidade de sucesso $P(y = 1) = \pi$ e de não sucesso $P(y = 0) = 1 - \pi$ temos que a respetiva função massa de probabilidade para cada observação é conhecida como:

$$2.4 \quad P(y = y|x) = p(x)^y (1 - p(x))^{1-y}$$

No modelo de regressão logística a relação entre a variável dependente, y , e independente, x , é estabelecida pela seguinte equação

$$2.5 \quad \pi(x) = E[Y|x] = \frac{e^{\beta_0 + x\beta}}{1 + e^{\beta_0 + x\beta}}$$

A função anterior só toma valores entre 0 e 1 (o que é espectável uma vez que se tratam de probabilidades) além disso pela derivação da mesma é possível concluir que só é crescente se $\beta_1 > 0$ Logo:

- $Y = 1 \rightarrow P(Y = 1) = \pi(x) = P(Y = 1|X = x) = \frac{e^{\beta_0 + x\beta}}{1 + e^{\beta_0 + x\beta}} \rightarrow \text{Sucesso}$
- $Y = 0 \rightarrow P(Y = 0) = 1 - \pi(x) = P(Y = 0|X = x) = \frac{1}{1 + e^{\beta_0 + x\beta}} \rightarrow \text{Insucesso}$

Conforme visto anteriormente a variável ε pode assumir qualquer valor, no entanto, para variáveis binárias o mesmo já não acontece:

- $Y = 1 \rightarrow \varepsilon = 1 - \pi(x)$
- $Y = 0 \rightarrow \varepsilon = -\pi(x)$

Portanto ε tem distribuição com média nula e variância $\pi(x)(1 - \pi(x))$.

Conforme foi dito anteriormente, considere-se que estamos perante uma variável resposta Y com distribuição *Bernoulli* $(1, p(x))$ e com a função de probabilidade citada anteriormente.

Uma vez conhecida a equação que estabelece a relação entre a variável dependente e independente para se poder ajustar o modelo é necessário estimar os valores de β_0 e β_1 até agora, parâmetros desconhecidos.

O método utilizado para a estimação dos mesmos é o método de Máxima Verosimilhança que, assumindo a independência das observações (x_i, y_i) , consiste na maximização da função de verosimilhança dada por:

$$2.6 \quad L(\beta) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n p(x_i)^{y_i} (1 - p(x_i))^{1-y_i}$$

A expressão da log-verosimilhança é dada como (onde $p(x_i) = p_i$):

$$2.7 \quad l(\beta) = \ln[L(\beta)] = \sum_{i=1}^n y_i \ln(p_i) + \sum_{i=1}^n (1 - y_i) \ln(1 - p_i)$$

A partir da igualdade a zero da derivada da log-verosimilhança aos respetivos parâmetros (β_0, β_1) é possível estimar os valores procurados:

$$2.8 \quad \frac{\partial \ln[L(\beta)]}{\partial \beta_0} = \sum_i \left[y_i - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right] = \sum_{i=1}^n [y_i - p_i]$$

$$2.9 \quad \frac{\partial \ln[L(\beta)]}{\partial \beta_1} = \sum_i \left[y_i x_i - x_i \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right] = \sum_{i=1}^n x_i [y_i - p_i]$$

As expressões anteriores resultam em equações não lineares o que impossibilita obter a sua solução de forma analítica. No entanto através de métodos numéricos como o de Newton-Raphson já é possível calcular a mesma.

Regressão Logística Múltipla

No modelo anterior, regressão logística simples, analisou-se a situação de termos apenas uma variável independente. Neste capítulo vai ser estudado o caso de existir mais do que uma variável, ou seja, como modelar uma relação entre uma variável binária dependente com mais que uma variável independente.

Apesar de se tratar de um método mais complexo que o anterior, não é nada mais do que uma generalização do mesmo mas aplicado a mais do que uma variável. De qualquer modo este método é uma análise simultânea entre variáveis independentes e dependente.

Assim, para aplicação da regressão múltipla pressupõe-se que variável dependente seja binária e as variáveis independentes sejam determinísticas.

Suponha-se agora que se trata de uma amostra com um conjunto p de variáveis independentes, $(x_1, x_2, \dots, x_n) = x^T$. Neste caso a probabilidade de um determinado evento Y ocorrer em função de x^T pelo modelo de regressão logística múltipla é dado do seguinte modo:

$$2.10 \quad P(x_1, x_2, \dots, x_n; \beta_1, \beta_2, \dots, \beta_n) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}$$

Onde, considerando x^T o vector de variáveis independentes e $\beta = (\beta_0, \beta_1, \dots, \beta_n)$ o vector de coeficientes que se pretende estimar. De uma forma matricial também se pode representar $P(x_1, x_2, \dots, x_n; \beta_0, \beta_1, \dots, \beta_n)$ do seguinte modo:

$$2.11 \quad p = P(x; \beta) = \frac{e^{x^T \beta}}{1 + e^{x^T \beta}} = \frac{1}{1 + e^{-x^T \beta}}$$

Analogamente à regressão logística simples para a estimação dos $n + 1$ parâmetros, até agora desconhecidos, é usual utilizar-se o método da Máxima Verosimilhança a partir de uma amostra $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$. Deste modo pressupõe-se a existência de p variáveis independentes do $p + 1$ vector (x_i, y_i) , a probabilidade de sucesso é dada por p_i e:

$$2.12 \quad L(y_1, y_2, \dots, y_p) = \prod_{i=1}^n P(x_{i1}, x_{i2}, \dots, x_{ip})^{y_i} (1 - P(x_{i1}, x_{i2}, \dots, x_{ip}))^{1-y_i} = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i}$$

Aplicando novamente a derivação da log verosimilhança é possível obter os valores procurados. No entanto mais uma vez as expressões anteriores resultariam em equações não lineares pelo que só é possível resolver por métodos numéricos como o de Newton-Raphson.

Escolha do Modelo

Teste de razão de Verosimilhança

Avaliar a qualidade do modelo estimado é um dos pontos mais importantes uma vez que analisa até que ponto o modelo é fidedigno na representação da realidade. Para isso é preciso testar a significância do mesmo que é feita através do teste de razão de verosimilhança (TRV) uma vez que os coeficientes foram determinados também pelo método da máxima verosimilhança.

O teste a aplicar vai testar se todos os coeficientes de regressão, $b = (\beta_0, \beta_1, \dots, \beta_p)$, são nulos com exceção de β_0 que corresponde ao termo constante.

Assim a hipótese a testar será: $H_0: \beta_1 = \beta_2 = \dots = \beta_p \quad Vs \quad H_1: \exists j: \beta_j \neq 0, \quad j = 1, \dots, p$

Estatística de teste: $\lambda = \frac{L_0}{L_1} \leq k$

Onde L_0 representa a função de verosimilhança já calculada sob a validade de H_0 , ou seja, $L_0 = L(y_0, \dots, y_n; \beta_0)$ e L_1 representa a verosimilhança calculado pelos estimadores de máxima verosimilhança sem nenhuma restrição, ou seja, $L_1 = L(y_0, \dots, y_n; \beta_1, \beta_2, \dots, \beta_p)$, ou seja:

- L_0 é a verosimilhança sem a variável, ou seja, a verosimilhança do modelo sem a covariável. (Verosimilhança do modelo com a interceção)
- L_1 é a verosimilhança com a variável, ou seja, a verosimilhança do modelo com a covariável. (Verosimilhança do modelo com as p variáveis explicativas)

Ajustando λ à expressão $-2\ln(\lambda)$ sabemos que, pelo teorema de Wilks, a mesma tem aproximadamente uma distribuição Qui-Quadrado com $p - 1$ graus de liberdade: $\lambda \rightarrow -2\ln(\lambda) \hat{\sim} \chi^2_{p-1}$

Ao se rejeitar a hipótese nula podemos concluir que existe pelo menos um dos coeficientes que é diferente de zero, significativo. Deste modo torna-se importante identificar qual dos coeficientes é que se trata. Para isso pode ser aplicado o teste de Wald ou Score que testa a significância dos parâmetros.

Teste Wald

O teste de Wald, descrito por Polit (1996) e Agresti (1990), é uma das possíveis formas de testar se os coeficientes associados a cada uma das variáveis explicativas tomam, ou não, o valor zero testando assim se as variáveis independentes são, ou não, significativas.

Ao contrário do teste da razão de Verosimilhança no teste de Wald é avaliado individualmente se cada variável independente é significativa ou não para a variável resposta, ou seja, testa se cada coeficiente é, ou não, igual a zero.

O teste de Wald é calculado a partir da relação entre a estimativa de máxima verosimilhança ($\hat{\beta}_i$) e a estimativa do seu erro padrão.

Hipóteses a testar: $H_0: \beta_i = 0 \quad Vs \quad H_1: \beta_i \neq 0, \quad i = 0, \dots, p$

Estatística de teste: $W_i = \frac{\hat{\beta}_i^2}{var(\hat{\beta}_i)} \cap \chi_1^2$

Caso H_0 seja rejeitado podemos concluir que $\hat{\beta}_i$ é diferente de zero e, por isso, x_i é significativa na interação com a variável dependente.

No entanto este teste pode ter algumas falhas. Deste modo é recomendado por alguns autores que seja utilizado o teste da razão de verosimilhança para testar se de facto o coeficiente não é significativo quando o teste de Wald não rejeita H_0 .

Seleção de Variáveis

No caso de existir mais do que uma variável independente a seleção das mesmas deve começar por uma análise individual de cada uma das variáveis para fazer uma espécie de pré-seleção e diminuir a entropia na análise múltipla (fase seguinte).

A significância de uma variável é medida pelo *p-value* de Wald, ou seja, quanto menor for este valor mais significativa é a variável. Deste modo, uma variável com um *p-value* inferior 0.28 será considerada para a análise múltipla no entanto caso seja superior o mesmo já não acontece.

De qualquer modo e, independentemente dos resultados anteriores, é importante conhecer o verdadeiro significado de cada variável uma vez que esta pode não ser considerada importante mas em termos práticos já ser relevante. Portanto, nestes casos, a variável é incluída para a análise múltipla apesar do resultado da “pré-selecção”.

Resumindo para a seleção de variáveis (no caso de existirem mais do que uma variável independente) numa primeira fase é feita a análise simples e antes de passar para a seleção múltipla deve ser feita uma interpretação prática dos resultados obtidos (das variáveis excluídas e das não excluídas).

Quando se minimiza o número de variáveis a incluir no modelo o mesmo torna-se numericamente mais estável e mais generalizado. Caso contrário, se existirem variáveis em excesso as mesmas podem aumentar os erros padrão.

A análise variável a variável resultou num conjunto de variáveis candidatas ao modelo final (conjunto \mathcal{C}). No entanto nem todas podem ser significativas. Para conseguir identificar as mesmas iremos utilizar um dos métodos: *Stepwise*, *Backward* e *Forward*.

Resumidamente, estes métodos são métodos iterativos que vão acrescentando ou retirando variáveis do modelo conforme o AIC com que o mesmo vai ficando, ou seja:

- *Forward*: este método além de, provavelmente, ser a técnica mais utilizada é considerado um método de inclusão progressiva uma vez que através do seu procedimento iterativo começa por formar um modelo com apenas uma variável e vai acrescentado variáveis ao modelo conforme a diminuição do valor do AIC até que não existam não variáveis significativas;

- *Backward*: enquanto que o método anterior começa sem nenhuma variável no modelo, neste caso acontece o contrário: o modelo inicia com todas as variáveis e em cada iteração vão sendo eliminadas até todas serem significativas. A eliminação das variáveis é novamente feita em função da diminuição do valor do AIC;
- *Stepwise*: Este método é uma adaptação do método *forward*. Este procedimento vai juntando progressivamente variáveis mas é acrescentada uma nova etapa em cada iteração. Depois de uma variável ser acrescentada é feito um novo teste onde se avalia a significância de todas as variáveis já incluídas e se retira aquelas que já não o forem. Este método é uma espécie de combinação com os dois métodos anteriores uma vez que se vão acrescentado variáveis mas sempre que se acrescenta é feita uma análise para confirmar que não existe nada “a mais”;

Um das vantagens destes métodos é serem relativamente acessíveis e de exigirem pouco esforço computacional. De qualquer modo é sempre preciso aplicar (sobre os modelos devolvidos) testes que avaliem a sua qualidade.

Uma vez aplicados os 3 métodos é usual utilizar AIC (Critério de Informação de Akaike) como critérios para selecionar o melhor modelo $AIC_p = -2 \log(L_p) + 2p$

Onde L_p é a função de verossimilhança do modelo e p é o número de variáveis explicativas consideradas no mesmo.

Qualidade do Modelo Final

A qualidade do modelo final é medida pelo ajuste que os valores estimados tomam em relação aos valores observados (quanto mais próximos forem melhor será o modelo). A análise a estas diferenças é conhecida pela análise dos resíduos para a validação da qualidade do modelo estimado.

Para este tipo de validação (qualidade do ajustamento do modelo) existem os seguintes métodos:

- Testes baseados em estatísticas dos resíduos avaliando o modelo de uma forma global:
 - Resíduos de Pearson
 - Resíduos da Deviance
- Representações gráficas dos valores dos resíduos: são comparados os resíduos para vários elementos

Resíduos de Pearson

Os resíduos de Pearson são medidos do seguinte modo:

$$2.13 \quad r_i = r(y_i, \hat{\pi}_i) = \frac{y_i - m_i \hat{\pi}_i}{\sqrt{m_i \hat{\pi}_i (1 - \hat{\pi}_i)}} \quad i = 1, 2, \dots, n$$

Sendo m_i é o número de observações com $x = x_i$, $\hat{\pi}_i$ é a probabilidade ajustada dos indivíduos em i .

Deste modo a estatística Qui-Quadrado de Pearson é dada por:

$$2.14 \quad R = \sum_{i=1}^n r(y_i, \hat{\pi}_i)^2 \cap \chi_{n-p-1}^2$$

Onde “ H_0 : O modelo ajustado aproxima-se dos valores observados”, ou seja, o modelo ajustado tem uma boa “qualidade”.

Resíduos de Deviance

Os resíduos de Deviance (soma dos quadrados dos resíduos SQ_e da regressão linear) são calculados do seguinte modo:

$$2.15 \quad d_i = d(y_i, x_i) = \pm \sqrt{2[y_i \ln\left(\frac{y_i}{m_i \hat{\pi}_i}\right) + (m_i - y_i) \ln\left(\frac{m_i - y_i}{m_i(1 - \hat{\pi}_i)}\right)]} \quad i = 1, 2, \dots, n$$

Sendo m_i é o número de observações com $x = x_i$, $\hat{\pi}_i$ é a probabilidade ajustada dos indivíduos em i .

Deste modo a estatística teste é dada por: $D = \sum_{i=1}^n d(y_i, \hat{\pi}_i)^2 \cap \chi_{n-p-1}^2$

Neste caso H_0 também significa “O modelo ajustado tem uma boa “qualidade””.

Teste de Hosmer-Lemeshow

Em 1989, Hosmer e Lemeshow, propuseram um novo teste para avaliar a qualidade de ajuste do modelo estimado (no caso de regressão logística).

O teste de Hosmer-Lemeshow é utilizado para testar o bom (ou não) ajuste do modelo, ou seja, através das distancias entre as probabilidades ajustadas e estimadas é avaliado a qualidade do mesmo.

Neste caso os dados são agrupados em g grupos segundo as suas probabilidades ajustadas. Os valores ajustados são ordenados por ordem crescente e sendo depois divididos em g grupos de tamanho aproximadamente igual. Hosmer e Lemeshow sugerem $g = 10$ no entanto não é consensual qual o valor que g deve tomar. No entanto o mesmo não deve ser inferior a 3 porque, nesse caso, não é se quer possível de calcular a estatística teste.

Antes de se iniciar o cálculo da estatística teste é sugerido que seja feita uma tabela de frequência esperada dentro de cada grupo (neste caso, aplicado para $g = 10$) uma vez que este teste vai comparar as frequências observadas e esperadas para avaliar se o modelo se ajusta bem à amostra:

Grupo - g	$Y = 0$		$Y = 1$		Total
	Observado	Esperado	Observado	Esperado	
1					
2					
...					
10					

Tendo a tabela acima preenchida torna-se mais fácil o calculo da estatística teste que é calculada do seguinte modo:

$$2.16 \quad C = \sum_{k=1}^g \frac{o_k - n_k \bar{\pi}_k}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)} \cap \chi_{g-2}^2$$

Onde:

n_k : é o número de indivíduos no k-ésimo grupo

o_k : soma dos valores da variável resposta no grupo k $\rightarrow o_k = \sum_{i=1}^{c_k} y_i$

c_k : número de covariáveis no grupo k

$\bar{\pi}_k$: média das probabilidades estimadas para o grupo k $\rightarrow \bar{\pi}_k = \sum_{i=1}^{c_k} \frac{m_i \hat{\pi}_i}{n_k}$

Sendo que a hipótese nula a testar é “O modelo representa bem os valores observados”

Curva ROC

Um conceito essencial para a interpretação da curva ROC é o ponto de corte. O ponto de corte corresponde a um valor de referência, x . Deste modo, as observações com probabilidades estimadas acima de x tomam o valor 1 e, abaixo, tomam o valor 0. São estes 0 e 1 se considera como os valores ajustados na curva ROC.

A curva ROC funciona como uma alternativa para avaliar o ajustamento do modelo é realizada também através da comparação entre os valores observados e os valores ajustados (0 e 1 anteriormente referidos). Quanto “menor” for a diferença entre estes dois valores melhor é o modelo.

A curva ROC é um dos principais métodos que permite testar a qualidade do comportamento de um modelo de regressão logística. Esta avaliação, pela curva ROC, é feita através da representação de um gráfico que vai relacionar a sensibilidade e especificidade para diferentes pontos de corte:

- *Sensibilidade*: define-se como a probabilidade de um resultado ser dado como positivo sendo que, na realidade, é, de facto, positivo;
- *Especificidade*: contrário à definição anterior, ou seja, define-se como a probabilidade de um resultado ser dado como negativo sendo que, na realidade, é, de facto, negativo;

Onde se considera a *sensibilidade* e a *especificidade* como:

$$2.17 \quad \text{Sensibilidade: } S = \frac{VP}{VP+FN}$$

$$2.18 \quad \text{Especificidade: } E = \frac{VN}{FP+VN}$$

Onde VP são os valores positivos previstos como tal, FP os valores negativos previstos como positivos, VN os valores negativos previstos assim mesmo e FN os valores positivos previstos como positivos.

Repare-se que a *sensibilidade* e a *especificidade* são independentes uma da outra uma vez que a *sensibilidade* usa os valores “positivos” enquanto que a *especificidade* usa os valores “negativos”.

Naturalmente que o objetivo é diminuir o mais possível o número de FP e FN, logo quanto menores forem estes valores melhor será o modelo. Estes valores dependem do ponto de corte escolhido, ou seja, se aumentarmos esse ponto FP diminui mas FN aumenta. Caso o ponto de corte seja diminuído acontece o inverso. Assim não é possível aumentar/diminuir FP sem que FN diminua/aumente.

Repare-se que não existe um ponto de corte ideal uma vez que o mesmo vai depender da natureza de cada problema. O valor de corte define-se como o ponto escolhido onde acima da qual o resultado é dado como positivo e abaixo do qual é dado como negativo. Logo para cada ponto de corte existe uma *especificidade* e uma *sensibilidade* diferente.

A curva ROC vai representar no ponto de corte a percentagem de VP Vs percentagem de FP. Uma “boa” curva ROC cresce rapidamente para 1 conforme o aumento do ponto de corte (o que leva a um afastamento da diagonal – a diagonal representa uma classificação aleatória) o ideal seria mesmo uma linha horizontal perto da unidade – o que é muito difícil de obter.

A curva ROC vai ser um instrumento fulcral para se poder definir qual o melhor valor para o corte.

Um outro aspeto muito importante na curva ROC é, precisamente, a área abaixo da curva uma vez que esta é utilizada para analisar as diferenças entre várias curvas ROC.

A área abaixo da curva devolve uma medida de discriminação, ou seja, um teste que não discrimina valores positivos de negativos teria uma área abaixo da curva ROC de cerca de 0.5 assim, quanto maior este valor maior seria a capacidade de discriminar os valores obtidos segundo estas duas hipóteses – positivo ou negativo. Analogamente para o caso da “curva ideal” (anteriormente definida) a sua área seria próxima de 1.

Assim é possível distinguir as seguintes áreas:

- $\text{Área}_{\text{abaixoROC}} = 0.5 \rightarrow$ não há discriminação;
- $0.7 \leq \text{Área}_{\text{abaixoROC}} \leq 0.8 \rightarrow$ discriminação tolerável;
- $0.8 < \text{Área}_{\text{abaixoROC}} \leq 0.9 \rightarrow$ discriminação boa;
- $\text{Área}_{\text{abaixoROC}} > 0.9 \rightarrow$ discriminação perfeita;

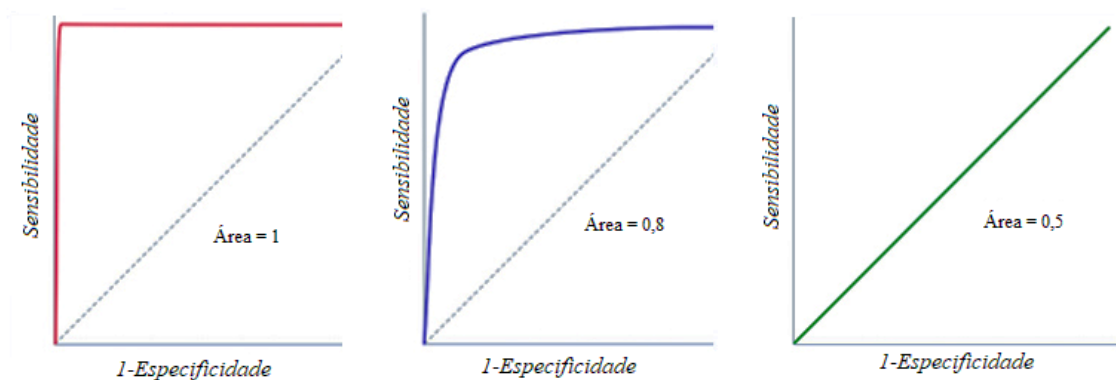


Figura 2.1 Três casos tipo da curva ROC

3. ANÁLISE GERAL À CARTEIRA

Estrutura do trabalho

Este capítulo está estruturado da seguinte maneira:

1. Tratamento da base de dados;
2. Análise descritiva à carteira;
3. Modelo de regressão logística para previsão de anulação das apólices na data corrente;
4. Análise da sinistralidade;
5. Aplicação dos resultados anteriores;

Relativamente à análise descritiva à carteira a mesma é feita em duas vertentes:

1. no que diz respeito ao perfil de tomador de seguro
2. relativamente à fidelização:
 - Crescimento da carteira;
 - Proporção de anulação de apólices;

Programa Utilizado

Como a amostra tinha muitas observações (apólices) com mais de 30 características cada tornou-se demasiado pesada para que fosse trabalhada em *Excel*. Assim, para otimizar, esta primeira fase do trabalho – tratamento de base de dados - foi utilizado o *Access*.

No entanto tanto para a análise descritiva como para a construção do modelo de regressão logística foi utilizado essencialmente o *R* mas também o *Excel*.

Base de Dados

Para a análise da carteira de apólices desta seguradora usou-se uma base de dados que inclui todas as apólices desde do início da companhia de seguros (2010).

Deste modo agruparam-se as variáveis em duas bases de dados distintas: apólices e sinistros.

Apólices (*ap*)

Relativamente às apólices consideraram-se as seguintes variáveis:

- Apólice: código identificador de cada contrato de seguro celebrado entre o tomador e a seguradora (*apolice*)
- Acta: código identificador por cada alteração feita na apólice (*acta*)
- Data de início: data de abertura da apólice (*datainicio*)
- Data de entrada: data de início da acta em referência (*dataentrada*)
- Data de anulação: data em que a apólice foi anulada (caso tenha sido). (*dataanulacao*)
- Fracionamento: período de pagamento do prémio (*fracc*)
 - Semestral (S)
 - Anual (A)
- Opção: tipo de cobertura da apólice segura (*opcao*)
 - Máximos
 - Mínimos
 - Médios

- Situação: estado da apólice (*situacao*)
 - Anulada
 - Em anulação
 - Normal
 - Suspensão de pagamento
- Categoria: tipo de viatura (*categoria*)
 - Ligeiro (LI)
 - Comercial (CM)
- Marca (da viatura segura) (*marca*)
- Data da primeira matrícula: informa sobre a idade da viatura segurada (*data1matricula*)
- Data da carta de condução (de cada condutor declarado): informa sobre os anos de experiencia (*datacartacond*)
- Data de nascimento (de cada condutor declarado): idade do tomador de seguro (*datanascond*)
- Sexo do tomador de seguro (*sexocondhab*)
- Localidade (*localidade*)
- Condutor declarado: indica se existe ou não condutor declarado (S/N) (*conddeclarado*)
- Número de condutores declarados (*numerocd*)
- Escalão *Bónus Males*: cada escalão corresponde a uma percentagem de agravamento/desconto sobre o prémio da apólice (*escalaobm*)
- Valor de *Bónus Males*: indica se o *BM* aplicado é de desconto, agravamento ou neutro
- Percentagem de *Bónus Males*: percentagem correspondente ao escalão *BM* (*percbm*)
- Zona: zona do tomador de seguro (*zona*):
 - Norte
 - Centro
 - Sul
 - Outros
- Cilindrada da viatura segura (*cilindrada*)
- Peso da viatura segura (*peso*)
- Serviço de reboque: viatura segura tem ou não serviço de reboque (*servicoreboque*)
- Assistência de viagem: tipo de assistência de viagem da apólice (*assistviagem*):
 - Base (comercial ou ligeiro)
 - VIP (comercial ou ligeiro)
- Franquia: percentagem da franquia caso exista (*franquia*)
- Colaborador: tomador de seguro é, ou não, colaborador na companhia de seguros (*colaborador*)
- Desconto ou agravamento: apólice tem desconto/agravamento (*descontoagravamento*)
- Valor em novo: valor da viatura segura em novo (*valornovo*)
- Valor reboque: valor do atrelado de reboque correspondente à marca/modelo da viatura segura (*valorreboque*)
- Valor tomador: valor que tomador de seguro sugere para capital seguro (pode variar 15% do proposto pela companhia de seguros e só é válida após aceitação da seguradora) (*valortomador*)
- Valor desvalorizado: valor da viatura segura actualizado (*valordesvalorizado*)
- Valor extra: valor correspondente a extras da viatura segura (*valorextra*)
- Custo da apólice segura durante o tempo seguro (*custoap*)
- Sinistros: número de sinistros da apólice (*sinistros*)

- Zona de circulação: zona de circulação da apólice segura – não tem de ser igual à localidade (*zonacircul*)
- Distrito: redução da variável *localidade* (*dist*)

Uma vez que a base de dados era muito pesada foram feitas algumas adaptações para facilitar a sua leitura:

- Diminuiu-se o número de níveis de algumas variáveis garantido que a perda de informação das mesmas não era significativa:
 - a variável *Marca e Modelo* foi categorizada apenas para a variável *Marca*. Esta redução levou a uma perda de informação importante uma vez que cada marca tem uma grande variabilidade de modelos. No entanto esta perda acaba por ser compensada pelas variáveis *cilindrada* e *peso* que traduzem algumas das mais importantes características das viaturas seguras)
- As variáveis *Data...* foram colocadas no mesmo formato: ano-mês-dia (AAAAMMDD)
- As células sem informação útil valem NA
- Para as apólices em vigor, como não existia data de anulação tomam a data de 1 de agosto de 2017 (dia em que os dados foram obtidos). No fundo é como se todas as apólices em vigor “congelassem” nesse mesmo dia.

Sinistros (*ps*)

Por fim relativamente aos sinistros considerou-se:

- Apólice: código identificador de cada contrato de seguro celebrado entre o tomador e a seguradora (*acta*)
- Sinistro: código identificador de cada sinistro regularizado pela companhia de seguros (*sinistro*)
- Tipo de sinistro: tipo de regularização de sinistro (*tipoids*)
 - IDS
 - Credor (2)
 - Devedor (1)
 - Misto (3)
 - Não IDS (via tradicional) (0)
- Data de sinistro: quando ocorreu o sinistro (*dataocorrencia*)
- Data de abertura: quando foi aberto o processo de sinistro (*dataparticipação*)
- Data do último pagamento (*datamovimento*)
- Código de risco: identificação do risco accionado em cada sinistro (note-se que cada cobertura pode ter riscos diferentes incluídos – por exemplo: a cobertura roubo e actos de vandalismo tem dois riscos incluídos roubo e actos de vandalismo) (*codrisco*)
- Provisão: custo final para cada código de cobertura
- Custo de processo de sinistro (*custops*)
- Custo por apólice: inclui o custo de todos os sinistros de cada apólice (*custoap*)

Analogamente às bases de dados anteriores esta foi igualmente adaptada para também facilitar a sua leitura:

- Criou-se a variável *codcobertura* onde se associa a cada risco o código de cobertura correspondente. Esta variável funciona como uma redução da variável *codrisco* uma vez que a mesma cobertura pode incluir mais que um risco
- As variáveis *Data...* estão no formato ano-mês-dia (AAAAMMDD)

- As células sem informação útil valem NA

Variáveis Extra

Uma vez trabalhadas as BD foram criadas variáveis que são obtidas a partir das BD originais, sendo elas:

- *Vigor*: identifica as apólices em vigor (ou não). Esta é uma variável binária sendo: 0 – “ap em vigor” e 1 – “ap anulada”;
- *Tempoap*: identifica o tempo de duração de cada ap conhecido até à data de anulação. Esta variável é obtida através do intervalo de tempo entre a data de anulação de ap e a data de início da mesma;

As variáveis anteriores vão ser melhor explicadas mais à frente no contexto da sua utilização.

Análise Geral à Carteira

A carteira de apólices em estudo tem cerca de 54% em vigor e com um tempo médio de duração próximo dos 2 anos.

Relativamente ao perfil do tomador de seguro trata-se, na sua maioria, de uma pessoa não colaboradora, com uma idade média de 47 anos, 24 anos de carta de condução e do sexo masculino (78%) (esta última percentagem pode ser alterada porque em 19% das apólices não é conhecida essa mesma informação).

As localidades mais comuns dos tomadores de seguro são Braga, Lisboa e Barcelos. Na maioria as localidades são do norte (com 63%). Relativamente à zona de circulação a maioria também corresponde à zona do Porto seguida pelas zonas de Lisboa e Braga com 38%, 14% e 13% respetivamente. No entanto esta informação pode ser alterada novamente, uma vez que em 19% das apólices nada se sabe sobre isso.

No que diz respeito à viatura segura, mais de 86% são viaturas ligeiras, com uma idade média de 13 anos, um peso de 1.691 kg, 1.601 cilindradas e sem serviço de reboque. A marca da viatura segura, torna-se difícil de avaliar, uma vez que existem mais de 62 marcas diferentes (o que leva a uma grande variabilidade das mesmas). No entanto a maioria das marcas são: Renault, Opel e Volkswagen. Expectavelmente quase 97% das viaturas são utilizadas para uso particular.

Nas características das apólices em si: 45% tem condutor declarado, 60% fazem o pagamento do prémio anualmente, 82% são apólices de mínimos, 16% de máximos e 2% de médios. Para a assistência em viagem: a maioria tem o serviço base para ligeiros. Note-se que este serviço corresponde ao tipo de opção da *ap* e categoria mais comum (opção de mínimos e categoria ligeiros).

Tendo em conta os sinistros ocorridos em cada apólice o custo médio é de 197,90€. No entanto este valor não deve ser interpretado desta forma porque não distingue os vários tipos de sinistro. Logicamente que sinistros com danos corporais terão custos muito superiores a sinistros apenas com danos materiais, apesar de estes últimos serem mais frequentes. Mesmo considerando apenas sinistros de danos materiais é expectável que exista uma grande variabilidade de custos (por exemplo: entre sinistros IDS e sinistros que são regularizados por via tradicional). De qualquer modo a sinistralidade vai ser abordada mais à frente.

Fidelização

O que torna o conceito de seguros funcional é a ideia de distribuição de risco, ou seja, aquando a ocorrência de um sinistro toda a carteira de apólices cobre o mesmo. Torna-se essencial que além de um bom saneamento exista também uma manutenção das apólices “boas”. Uma vez que o saneamento elimina as *ap* que trazem excesso de risco para a seguradora e a manutenção das “boas *ap*” alivia o risco que recai sobre a mesma.

É com o intuito de manter pelo máximo tempo possível as apólices “boas” que surge a preocupação de fidelizar as mesmas. Esta fidelização pode ser feita de muitas maneiras como: aplicação de descontos, oferta de experiências (massagens, desportos radicais...), campanhas de marketing...

Constituição da Carteira

A fidelização de bons clientes é, de facto, um dos fatores que mais estabilidade e segurança oferece à companhia de seguros. Com o intuito de analisar a fidelização dos clientes começou-se com uma análise ao tempo de apólice segura. Para isso a variável em estudo é *tempoap* que foi calculada do seguinte modo:

$$3.1 \quad \textit{tempoap} = \textit{datainicio}_i - \textit{dataanulação}_i \quad i = 1, \dots (\textit{apólice})$$

No que diz respeito às apólices que estão em vigor foi considerado que todas elas “congelaram” em 2017-08-01. Dado que não é conhecida nenhuma informação sobre a continuidade ou não das mesmas no período posterior ao indicado.

Numa fase inicial detetou-se que na carteira existem cerca de 8,44% de *ap* que têm duração de 0 dias, isto significa que estas apólices foram criadas mas que o primeiro prémio não foi pago sendo automaticamente anuladas. No entanto apesar de não existirem custos com sinistros existiram custos administrativos, sem nenhum pagamento recebido, pelo que se optou por incluir estas mesmas *ap*.

De qualquer modo apesar de os custos destas *ap* não serem significativos, em percentagens mais elevadas podem de facto ter consequências mais pesadas. Seria interessante “medir” a partir de que ponto estas apólices afetam a seguradora.

Protegendo a seguradora com medidas como: cobrança de um valor inicial apenas pela criação da apólice que depois seria (ou não) amortizado no valor do primeiro prémio a pagar. No entanto este tema não vai ser abordado uma vez que o que se procura é avaliar a fidelidade das apólices já criadas.

Assim, na nossa carteira de apólices temos:

- Valores extremos: o valor mínimo é de 0 dias pelos motivos anteriormente explicados e o máximo é de 7 anos o que também é esperado uma vez que se trata de uma seguradora com 7 anos. Portanto, existe pelo menos uma *ap* que acompanha a seguradora desde o seu início.
- Uma vez que a amostra tem apenas 7 anos a variabilidade da mesma torna-se difícil de medir.
- Os dados são assimetricamente (assimetria positiva) - distribuídos uma vez que a mediana se encontra perto do primeiro quartil.
- Uma vez que a média e a mediana, em valores aproximados, são iguais pode-se concluir que a amostra tem pouca variabilidade.
- Que com base na média quer com base na medianas *ap* têm uma duração média perto dos 2 anos. Onde 75% dura 3 anos e 25% um ano.

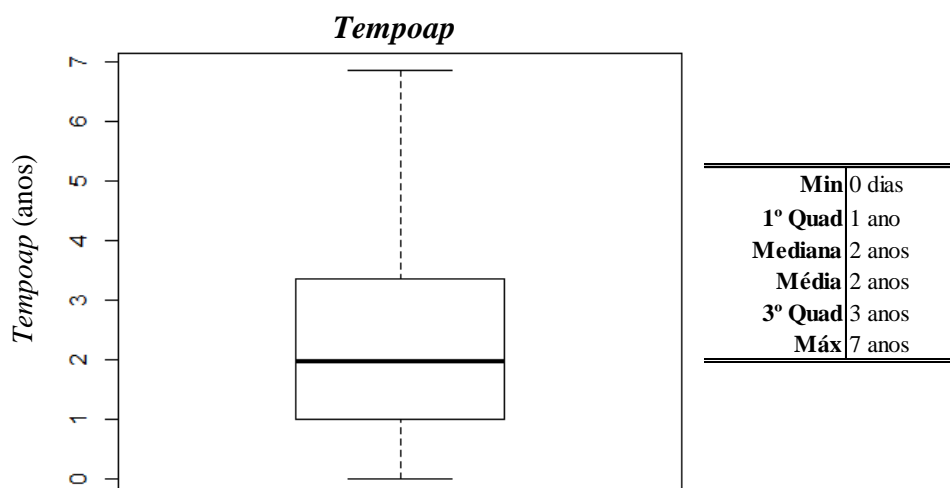


Figura 3.1 Tempo de *ap* segura (em geral)

A seguradora em causa tem com poucos anos de existência, 7, mas a maioria das *ap* não chegam a metade desse tempo. O que pode ser um sinal de alerta em termos de retenção de clientes.

Apesar de não existir nenhuma referência para a duração “ideal” das *ap*, 2 anos parece pouco.

Apólices de Colaborador Vs Não Colaborador

Uma vez que a duração das *ap* pareceu ser tão baixa esta situação foi analisada mais a fundo. Procurou-se avaliar se seria um “problema” geral ou apenas de um grupo específico da carteira.

Analisando as variáveis disponíveis optou-se por distinguir *ap* de *colaboradores* e de *não colaboradores*, uma vez que são aplicadas tarifas diferentes a estes dois grupos. As *ap* de *colaboradores* têm mais descontos o que pode incentivar à manutenção da *ap* na seguradora.

Na carteira usada como amostra existem cerca de 5% de apólices pertencentes a *colaboradores*.

Conforme esperado foram obtidos os seguintes resultados (nos gráficos abaixo ocultaram-se os *outliers*):

- Apenas apólices de *colaboradores*:
 - Valores extremos: o valor mínimo mantém-se em 0 dias pelos mesmos motivos assim como o máximo se mantém em 7 anos
 - O valor da média e da mediana são bastante próximos (em valor absoluto) com base na média e na mediana as *ap* de *colaboradores* têm uma duração de 4 anos
- Apenas apólices de *não colaboradores*:
 - Valores extremos: conforme conclusões anteriores trata-se de um mínimo de 0 dias e um máximo de 7 anos
 - existe uma variabilidade maior pode-se concluir que as *ap* de *não colaboradores* têm uma duração de 2 anos
 - os dados são assimetricamente positivos distribuídos uma vez que a mediana se encontra relativamente perto do primeiro quartil

- Comparando as *ap* destes dois grupos pode-se ver que: quer o 1º quartil, 2º ou 3º assim como na média as *ap* de *colaboradores* têm um *tempoap* acima das *ap* de *não colaboradores*.

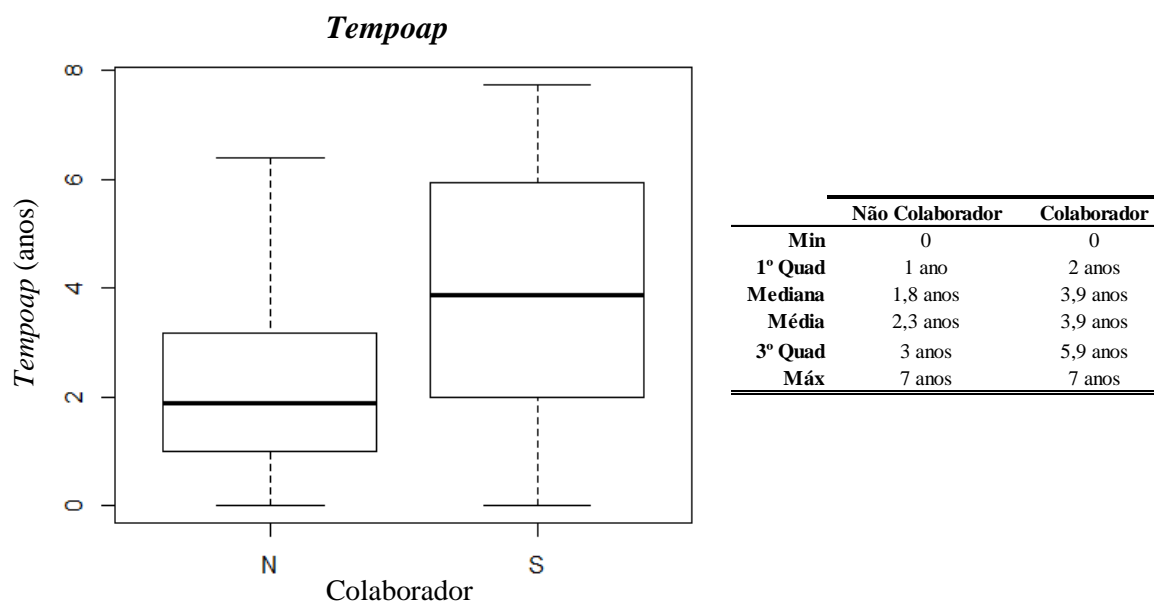


Figura 3.2 Tempo de *ap* segura entre *ap* de colaborador e não colaborador

Para confirmar a suspeita a cima aplicou-se o teste de Wilconx-Mann-Whitney onde se verificou que o *tempoap* é significativamente superior para *ap* de *colaboradores* ($p - value < 2.2e - 16$).

Conclui-se que as diferentes condições das *ap* de *colaboradores* (essencialmente descontos) são significativas para a durabilidade da mesma.

Relembre-se, que o facto de não ser colaborador não significa que a *ap* não tenha descontos. Existem *BonusMales* ou outros descontos que podem ser aplicados independentemente do tipo de *ap*. Este facto pode justificar a existência de uma maior variabilidade para *ap* de *não colaboradores*. Uma vez que, já se provou que os mesmos influenciam na duração.

Devido ao alerta anterior (falta de fidelização dos clientes) analisou-se a distribuição por *tempoap*. Onde, mais uma vez, se comprovou que a fidelização dos clientes é um problema nesta amostra:

- na carteira de *ap* em geral 27 % das mesmas duraram mais de 3 anos. E a partir das *ap* com mais de 5 anos a diferença entre anos consecutivos vai diminuindo significativamente
- *Colaboradores*: mais de 43% das *ap* dura mais de 4 anos e 6% das *ap* dura 7 anos
- *Não colaboradores*: 25.5% das *ap* dura mais de 3 anos e somente 0,4% dura 7 anos
- Conclusão: uma vez que apenas 5% da carteira é constituída por *ap* de *colaboradores* estas não têm um impacto significativo. Assim a variação da carteira total é muito semelhante à variação das *ap* de *não colaboradores*. Nota-se uma diferença significativa entre *ap* de *colaboradores* e de *não colaboradores*

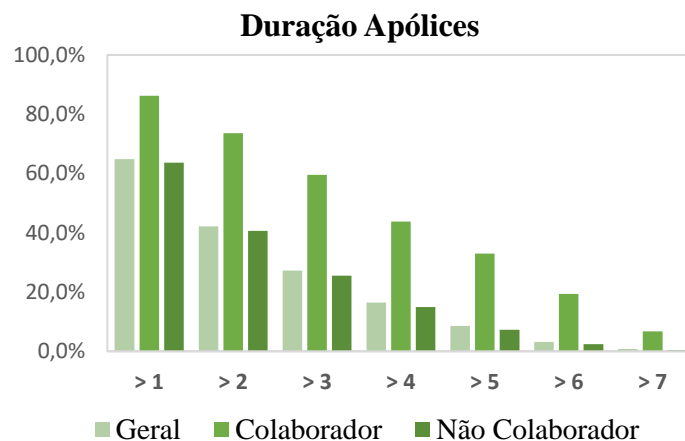


Figura 3.3 Percentagem de *ap* (geral, colaborador e não colaborador) por tempo seguro

Tabela 3.1 Percentagem de *ap* (geral, colaborador e não colaborador) por tempo seguro

	> 1	> 2	> 3	> 4	> 5	> 6	> 7
Geral	64,8%	42,2%	27,2%	16,4%	8,6%	3,2%	0,8%
Colaborador	86,2%	73,6%	59,5%	43,8%	33%	19,4%	6,7%
Não Colaborador	63,7%	40,6%	25,5%	14,9%	7,3%	2,4%	0,4%

Pelo gráfico quase metade das *ap* dura no máximo 2 anos das quais quase 40% dura no máximo 1. Analogamente nos *não colaboradores* mais de 55% não dura mais de 2 anos, 30% das quais não dura mais de 1 ano. O mesmo já não acontece com *ap* de *colaboradores* que além de terem uma distribuição mais uniforme entre cada classe, 59% dura no máximo 4 anos dos quais apenas 14% dura menos de 1 ano.

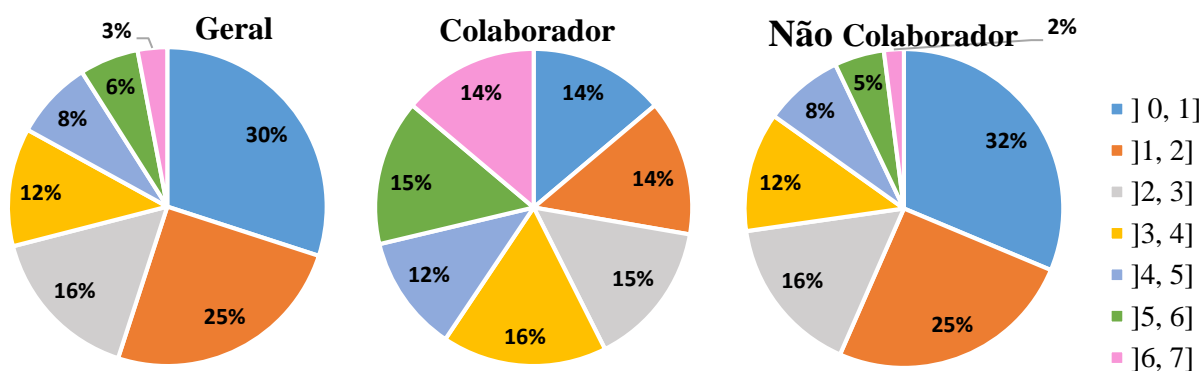


Figura 3.4 Percentagem de *ap* (geral, colaborador e não colaborador) por intervalo de tempo

Tabela 3.2 Distribuição de apólice segura por intervalo de tempo (geral, colaborador e não colaborador)

] 0, 1]] 1, 2]] 2, 3]] 3, 4]] 4, 5]] 5, 6]] 6, 7]
Geral	30%	25%	16%	12%	8%	6%	3%
Colaborador	14%	14%	15%	16%	12%	15%	14%
Não Colaborador	32%	25%	16%	12%	8%	5%	2%

Prova-se novamente que a aplicação de descontos leva a um aumento significativo de *tempoap*.

Taxa de Anulação

Uma vez conhecida a maneira como as *ap* se comportam em termos de duração, é importante conhecer as suas taxas de anulação. É esta variável que determina a redução da carteira de *ap* em estudo.

Nas taxas de anulação tem-se em consideração o número de *ap* anuladas e a quantidade de *ap* existentes. Uma vez que o importante não é o número de *ap* anuladas mas a sua percentagem.

$$3.2 \text{ Taxa anulação}_i = \frac{\text{apólices anuladas}_i}{\text{apólices criadas}_i + \text{apólices existentes}_{i-1}} \quad i = 2010, \dots, 2017 \text{ (anos)}$$

Como foi visto anteriormente, existe uma alguma variabilidade na variável *tempoap*. Assim, com o intuito de diminuir a mesma criaram-se as seguintes categorias:

- Independentemente dos *tempoap* - geral
- *Ap* com menos de 1 ano (< 1)
- *Ap* com pelo menos 1 ano (≥ 1)
- *Ap* com mais de 2 anos (média do *tempoap*) (> 2) – valor médio de *tempoap* da carteira

Relembre-se que se considerou *tempoap* como o intervalo de tempo entre a data de criação e a data de anulação de cada *ap*, ou seja, o tempo em que a *ap* esteve em vigor.

Também seria interessante analisar *ap* com mais de 3 anos de *tempoap* (3º quartil). No entanto, como a amostra tem só 7 anos dos quais os 2 primeiros dizem respeito à criação da seguradora e o último ano (2017) não está completo, restariam apenas 4 anos “úteis” para analisar. Logo, como não existe informação suficiente para se tirar conclusões ou prever qual a evolução para os anos seguintes, optou-se por não se analisar esta categoria em específico.

Relativamente à taxa de anulação podemos concluir que:

- A taxa de anulação para *ap* com menos de 1 ano é a que tem valores mais elevados. Este facto já é esperado uma vez que são nestas *ap* que se encontram situações como: burla, excesso de sinistralidade, *ap* criadas e anuladas no mesmo dia...
De qualquer modo, esta classe é um sinal de alerta pois apesar dos casos identificados acima, esta taxa é aparentemente elevada comparando com as outras categorias. Este é um sinal de que a seguradora não está a fidelizar os clientes logo nos primeiros anos.
- De um modo geral as taxas de anulação foram aumentando até 2014, o que é normal uma vez que a companhia de seguros começou em 2010 e, portanto, houve uma forte campanha inicial.
- 2013 e 2014 foram dois anos de crise pelo que de 2012 para 2013 houve um grande aumento da taxa de anulação atingindo o seu valor máximo em 2014.
- O ano de 2014 foi o ano em que mais se sentiu o efeito da crise.
- Depois da crise as taxas baixaram um pouco mas mantiveram-se relativamente constantes. O que pode significar que não houve ou não teve efeito qualquer esforço feito para recuperar os efeitos dos anos anteriores. Facto este comprovado também pelas *ap* com mais de 2 anos que continuaram a subir a respetiva taxa de anulação.
- Conforme explicado anteriormente no que diz respeito às *ap* em vigor estas foram “congeladas” portanto o seu tempo de apólice segura pode ser superior ao considerado. Deste modo nas categorias: superior a 2 anos ou superior a 1 ano os últimos anos podem ter resultados ligeiramente alterados face à realidade.

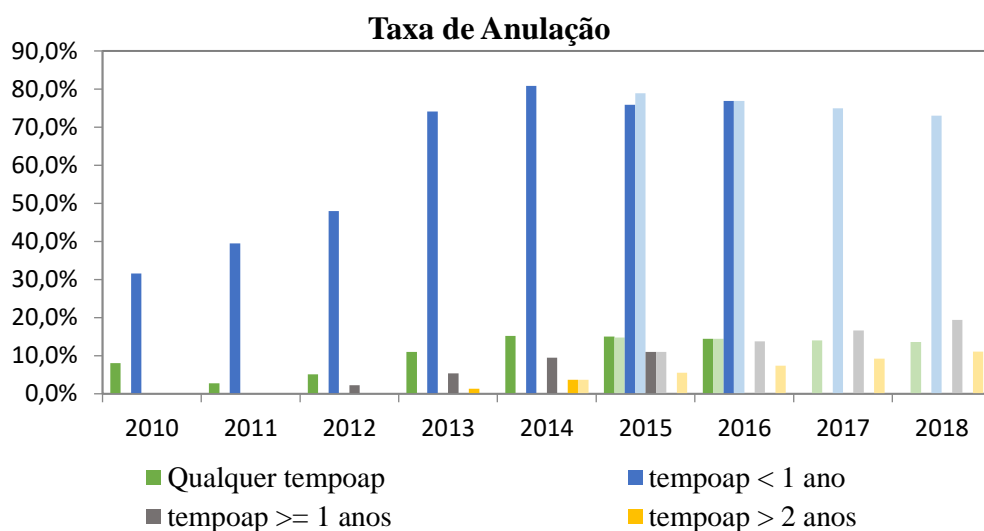


Figura 3.6 Taxa de anulação anual

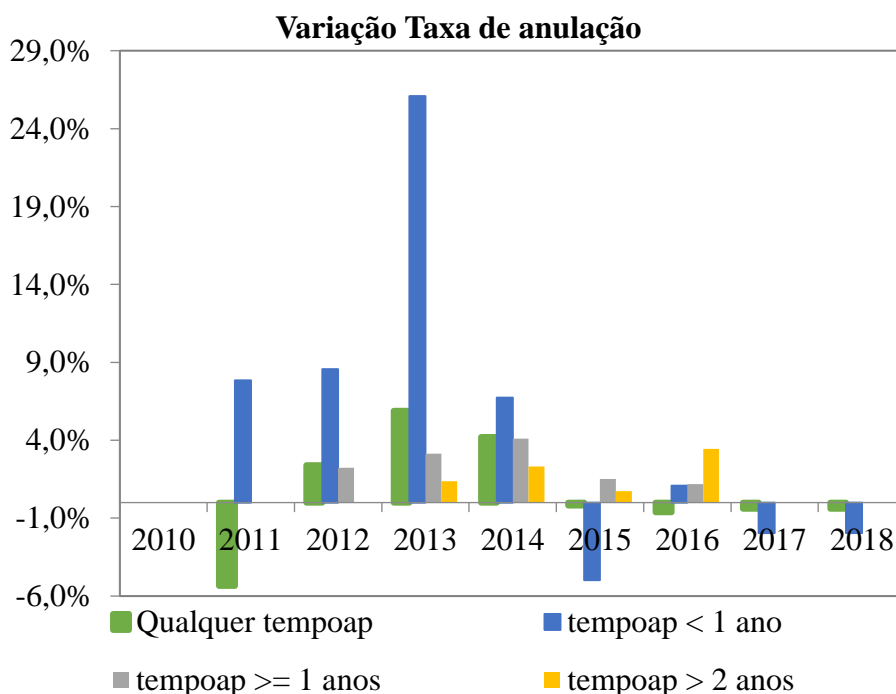


Figura 3.5 Variação da taxa anual entre anos consecutivos I

Na previsão da evolução da carteira na taxa de anulação supôs-se o seguinte (barras com tons mais claros no gráfico da taxa de anulação anual):

- Uma vez que a seguradora se iniciou em 2010 considerou-se que os anos de 2010, 2011 e 2012 foram anos de campanha, crescimento. Portanto não se considerou como “anos exemplo” para uma possível tendência;
- 2013 e 2014 foram anos de crise logo também não se considerou como uma tendência;
- Deste modo, como temos uma amostra de apenas 7 anos em que o último não está completo considerou-se que:
 - Para as classes de *Qualquer tempoap* e *tempoap < 1 ano* os anos de 2015 e 2016 como anos úteis para uma previsão da taxa de anulação;
 - Para as classes *tempoap >= 1 ano* e *tempoap > 2 anos* já se considerou os anos de 2014-2015 e 2013-2014 respetivamente, pois estas categorias foram “cortadas” pelo “con-

gelamento”. No entanto os anos considerados foram de crise por isso as previsões destas categorias são pouco fidedignas. Principalmente para a classe *tempoap* > 2 anos que foi a mais “afetada”;

Repare-se que a previsão foi calculada a partir da média dos anos anteriores.

Uma vez que o ano de 2017 não está completo e se teve de “congelar” as *ap* em vigor fez-se a previsão não só para 2018 mas também para 2017.

Pode-se concluir que de um modo geral vai haver uma estabilização na taxa de anulação, com uma ligeira tendência a diminuir, excepto nas classes *tempoap* >= 1 ano e *tempoap* > 2 anos. No entanto estas duas últimas classes são pouco fidedignas pelos motivos anteriormente explicados.

Pelo gráfico seguinte é possível ter uma noção de como a taxa de anulação foi variando entre anos consecutivos:

- Existência de um pico em 2013 que advém da crise – subida significativa 2012-2013
- Nos anos de 2014-2015 e 2015-2016 há uma estabilidade entre as taxas (variação das mesmas é muito baixa). Apenas as *ap* com menos de 1 ano têm uma descida relevante.
- Para a *categoria* > 2 o gráfico termina no ano 2014 assim como para a *categoria* ≥ 1 devido às *ap* congeladas não permitirem tirar conclusões fidedignas

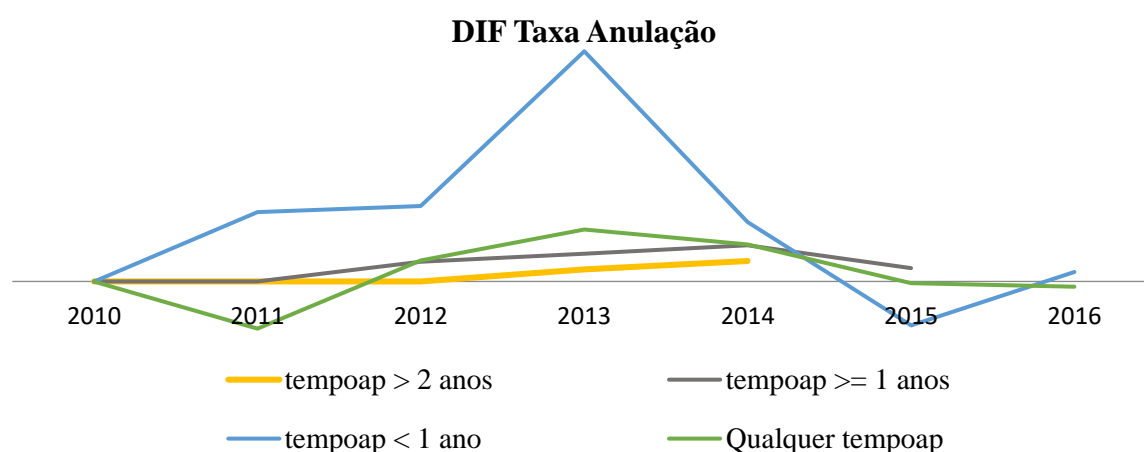


Figura 3.7 Variação da taxa anual entre anos consecutivos II

Apólices Criadas

O que leva ao crescimento de uma carteira é, a diminuição da taxa de anulação, mas também a criação de novas *ap*. A não anulação leva à estabilização da carteira, mas a criação de novas *ap* leva ao crescimento da mesma.

Enquanto que na taxa de anulação se analisa a fidelização dos clientes, a análise das *ap* criadas analisa a “atração” que a companhia tem no mercado.

Obviamente, que mais importante que ter uma carteira em crescimento é ter uma carteira de qualidade. Pressupõe-se que é aplicado um bom saneamento, pelo que a preocupação da fidelização recai apenas sobre as apólices “boas”.

Aplicando as mesmas classes que na análise da taxa de anulação, concluiu-se que:

- a criação de *ap* tem vindo a aumentar – independentemente da categoria
- Existe uma quantidade maior de *ap* criadas para mais de um 1 e menos para *ap* com menos de

1 ano. O que é positivo pois demonstra que são angariadas *ap* de melhor “qualidade”. No entanto a categoria para mais de 1 ano inclui um maior intervalo de tempo o que pode justificar a situação anterior e não ser um sinal assim tão positivo

- A relação entre a quantidade de *ap* angariadas de cada categoria se mantém constante ao longo dos anos. Pode-se deduzir que poderá não ter existido nenhuma campanha direcionada em específico para nenhuma delas
- Nos anos de 2013 e 2014 não houve o mesmo tipo de crescimento que os anos anteriores. Facto este que se deve às consequências da crise.

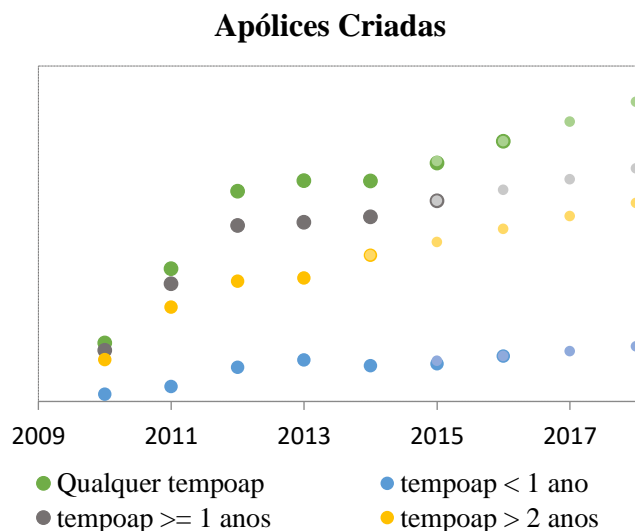


Figura 3.8 Quantidade de *ap* criadas anualmente

Analogamente às previsões feitas para a taxa de anulação, o número previsto para as *ap* criadas nos anos de 2017 e 2018 foi feito do mesmo modo e com as mesmas suposições. Mais uma vez, os pontos claros no gráfico da figura 3.8 dizem respeito às previsões.

Deste modo prevê-se um crescimento contínuo na criação de *ap* em geral (isto é, independentemente, da classe aplicada).

Para analisar qual a intensidade com que a criação de *ap* varia entre anos consecutivos, segue o gráfico abaixo:

- o grande pico de crescimento deu-se com a abertura da companhia de seguros
- a grande descida deu-se devido à crise
- pelos mesmos motivos que na taxa de anulação ignorou-se os anos de 2015 e 2016 para a categoria de superiores a 2 anos e o ano de 2016 para a categoria superior a 1 ano

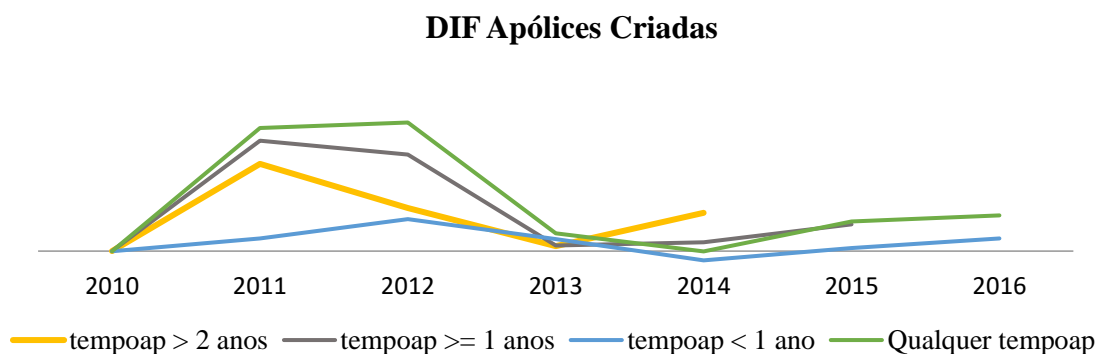


Figura 3.9 Variação do número de apólices criadas anualmente

Crescimento Carteira de Apólices

Nas páginas anteriores foi analisado individualmente a taxa de anulação e a criação de apólices. No entanto é na interação entre estes dois fatores que se devolve a verdadeira evolução da carteira de *ap*. Esta relação é bastante importante, porque é a partir daqui que a seguradora avalia o seu crescimento, para depois analisar (individualmente) o que de facto pode estar a prejudicar o mesmo.

A análise do crescimento da carteira é feita a partir da quantidade de *ap* em vigor. Deste modo fez-se o seguinte cálculo:

$$3.3 \quad ApolicesVigor_i = ApolicesCriadas_i + ApólicesVigor_{i-1} - ApolicesAnuladas_i \\ i = 2010, \dots, 2017 \text{ (anos)}$$

Analogamente aos capítulos anteriores, esta análise foi feita desde 2010 até 2016, deixando o ano de 2017 e 2018 para o cálculo de previsão. Dado que são anos onde não existe informação ou a informação que existe não está completa (como é o caso de 2017).

Verificou-se o crescimento também em quatro níveis (coerentemente com as análises anteriores). De onde se concluiu:

- Qualquer que seja a classe o seu crescimento é constante ao longo dos anos. Este facto que é comprovado pelo gráfico e pela pouca variação entre as *ap* anuladas e as criadas
- Apesar de não se notar nenhum pico significativo no crescimento da carteira, nos anos de 2013 e 2014 houve um crescimento menor.
- A relação entre as classes mantém-se a mesma entre todos os anos
- Relativamente à previsão espera-se que a evolução se mantenha a mesma que nos anos anteriores (pontos a mais claros no gráfico da figura 3.10)

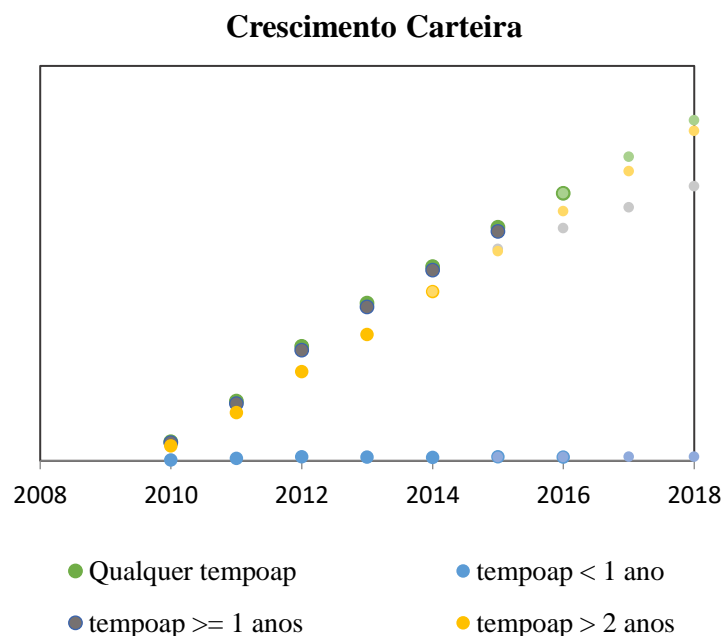


Figura 3.10 Crescimento da carteira de *ap*

DIF Crescimento de Carteira

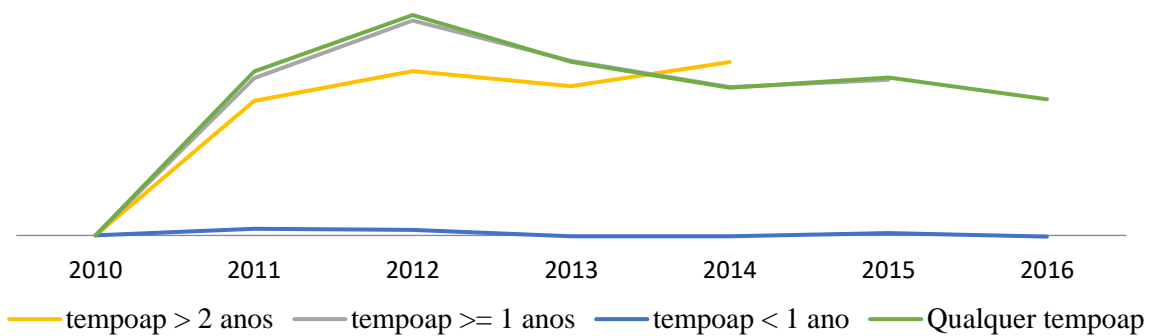


Figura 3.11 Variação do crescimento da carteira de *ap*

Reunindo as quatro categorias, podemos ver que apenas a categoria com *tempoap* < 1 (a azul) tem uma quantidade de *ap* criadas muito superior à quantidade das *ap* que existem. Aparentemente este facto não faria sentido uma vez que a quantidade de *ap* em carteira corresponde às *ap* criadas mais as que estão em vigor do ano anterior. Logo seria expectável que o total de *ap* em cada ano fosse superior ao número de *ap* criadas nesse mesmo ano, o que não acontece. No entanto a taxa de anulação é tão elevada que leva a que exista esta diferença tão grande e contrária às outras categorias. Novamente se verifica que a classe *tempoap* < 1 além de corresponder a 30% das *ap* tem taxas de anulação muito elevadas.

Conforme conclusões anteriores podemos afirmar que a carteira de *ap* em análise tem um grande problema de falta de fidelização dos segurados devido às taxas de anulação tão elevadas.

De qualquer modo já foi também provado que as condições aplicadas a colaboradores incentivam ao aumento da duração das mesmas.

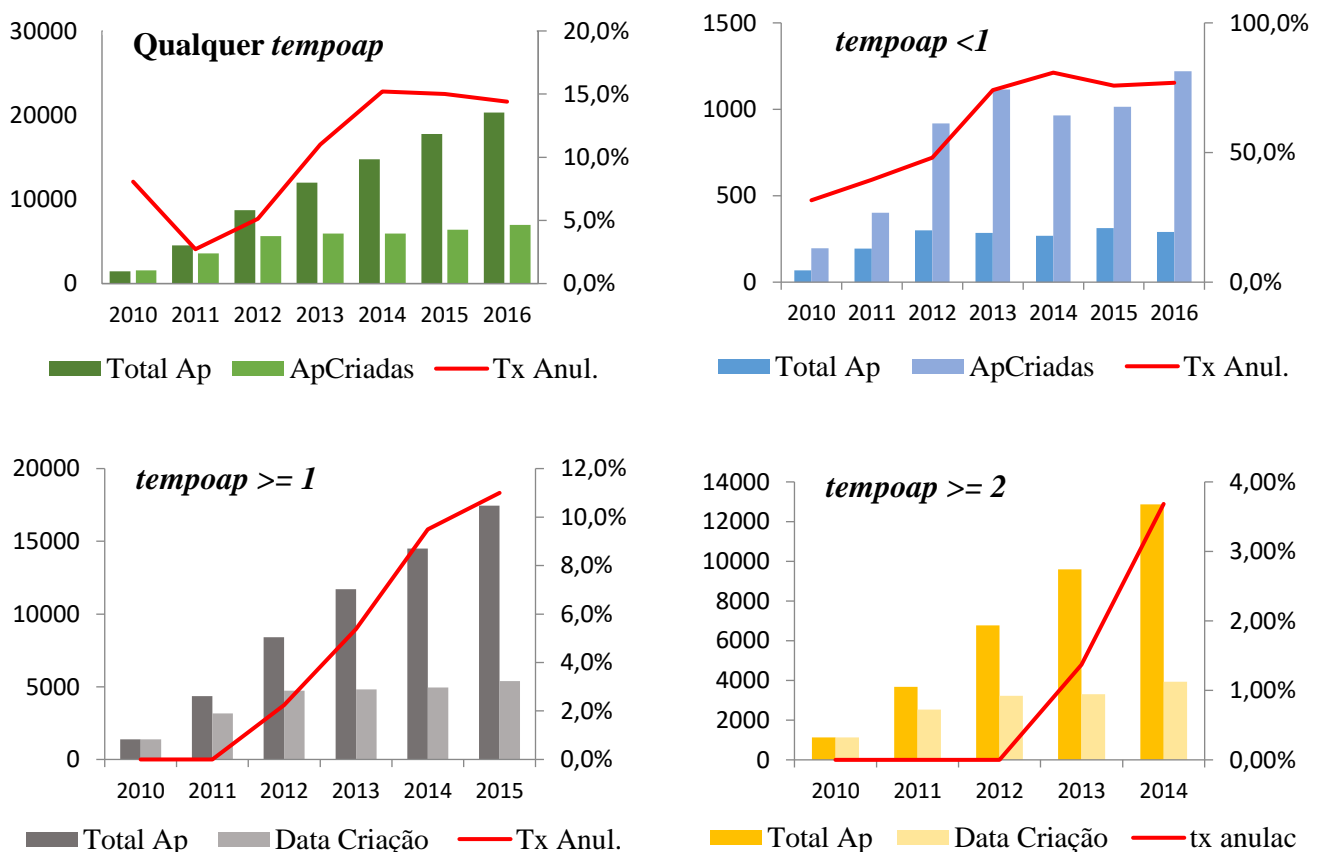


Figura 3.12 Comparação entre classes da taxa de anulação, criação de apólices e crescimento da carteira

Proporção de Apólices Anuladas

Uma vez que se detetou a existência de falta de fidelização dos clientes, esta questão foi abordada mais profundamente. Procurou-se então acompanhar as *ap* deste a sua data de criação.

Considere-se que se ignorou novamente o ano de 2017 por não estar completo.

Numa fase inicial começou-se por se identificar a quantidade de *ap* anuladas tendo em conta o seu ano de criação. Com base nesta informação é possível “acompanhar” cada apólice tendo em conta o ano em que se iniciou. Com esta informação calculou-se a proporção de *ap* anuladas que vai permitir tirar conclusões no que diz respeito à duração das mesmas.

Na tabela abaixo é visível que com menos de 3 anos mais de metade das *ap* criadas foram anuladas – o que no capítulo anterior já tinha sido detetado. Note-se que em 6 anos restariam apenas pouco mais de 30% das *ap*. No entanto este valor diz apenas respeito às *ap* com data de criação de 2010.

Tabela 3.3 Proporção de apólices anuladas por ano de criação - acumulada

		Ano Anulação						
Ano Criação		2010	2011	2012	2013	2014	2015	2016
	2010	8,07%	24,73%	36,62%	45,90%	52,13%	58,04%	63,64%
	2011		5,81%	24,72%	38,44%	49,10%	56,48%	62,15%
	2012			10,99%	31,97%	45,86%	55,82%	62,77%
	2013				14,00%	33,87%	46,30%	56,86%
	2014					11,77%	29,73%	43,89%
	2015						10,95%	31,82%
	2016							13,32%
Média		10,70%	29,47%	42,22%	51,92%	57,13%	60,10%	63,64%

Conforme gráficos abaixo é possível concluir o seguinte:

- Os anos de criação com menor taxa de anulação foram os primeiros anos da companhia (2010 e 2011) contrariamente aos anos de 2012 e 2013 – que coincidem com a época de crise em Portugal
- apenas de se tratar de uma amostra de apenas 7 anos é visível que existe um ligeiro pico no 1 ano das *ap*. No entanto, com o passar dos anos, começa-se a notar uma estabilização na anulação.

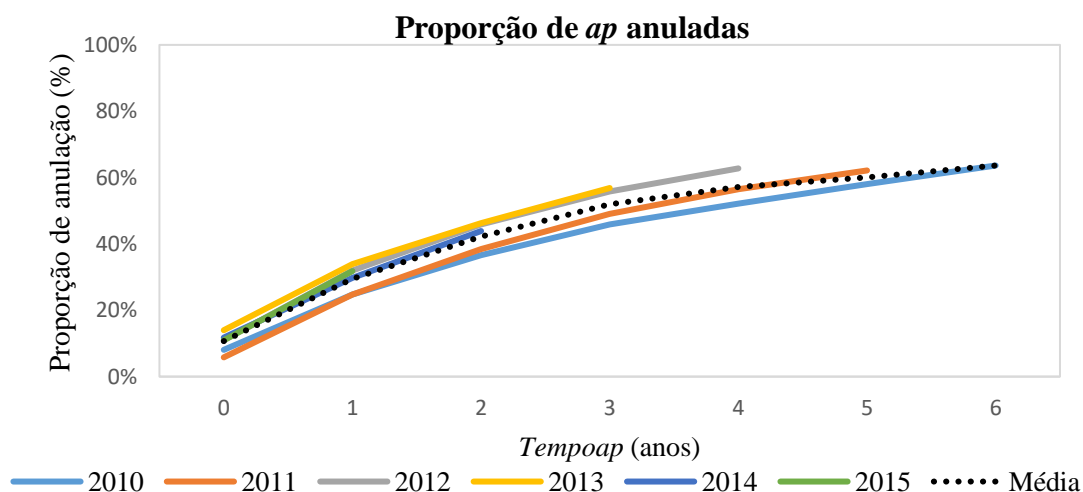


Figura 3.13 Proporção de *ap* anuladas

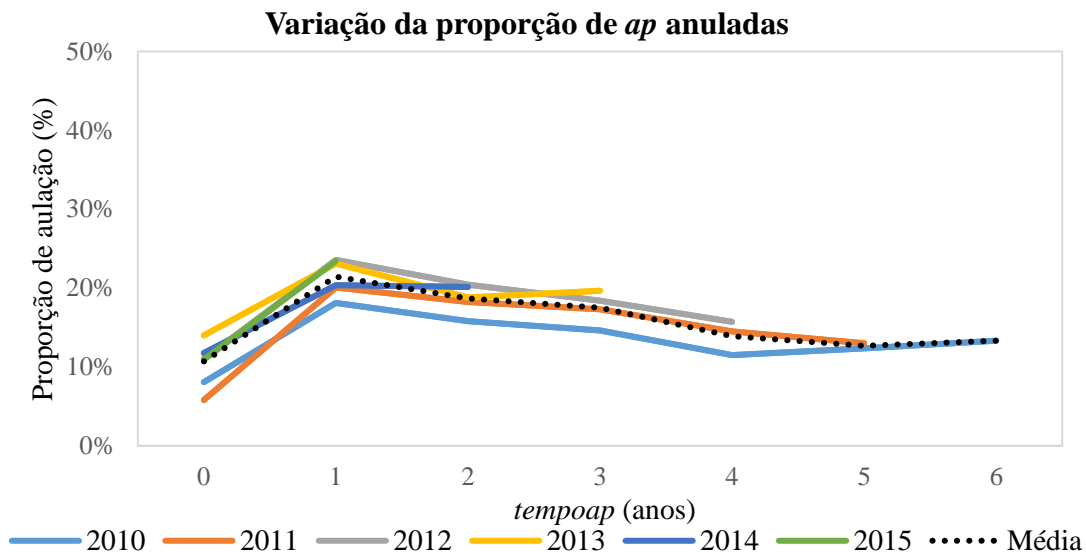


Figura 3.14 Variação da proporção de *ap* anuladas

O último facto anteriormente identificado pode ser muito importante para o modelo que se pretende criar. Uma vez que se torna expectável que duas *ap* com características iguais mas *tempoap* diferentes poderão ter probabilidades de anulação também diferentes. De qualquer modo seria preciso mais tempo de amostra para comprovar com maior segurança a dedução acima.

Assim, tendo em conta a estabilização dos clientes ao longo do tempo e o pico no primeiro ano de *ap* segura é possível identificar pelo menos duas fases de *tempoap*:

- fase (1): fase crítica uma vez que qualquer que seja a data de criação nota-se que grande parte das *ap* é anulada nesta fase;
- Fase (2): fase de estabilidade uma vez que se mantêm as apólices “boas” (as menos boas são saneadas). Nesta fase começam a surgir as apólices “fiéis”, ou seja, mantêm-se na companhia precisamente pela companhia e não por outra questão como o do custo;

É precisamente na “área” da fase 1 e da fase 2 que se vai refletir a retenção dos clientes. Ou seja, quanto maior a fase 1 menor será a fidelização dos mesmos. Deste modo o objetivo será começar a fase 2 o mais “cedo” possível.

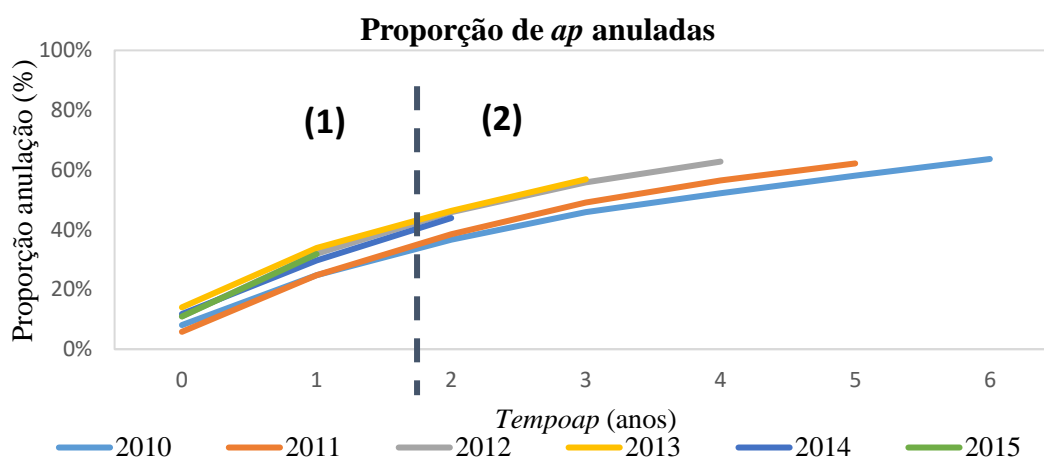


Figura 3.15 Fase (1) e fase (2) na anulação de *ap*

Não acumulando as *ap* anuladas é possível comprovar novamente que as maiores percentagens de anulações ocorrem durante no primeiro ano, onde cerca de um quarto destas apólices se anula.

Por esta análise das proporções de anulação é possível deduzir que a probabilidade de uma *ap* ser anulada vai depender também do respetivo *tempoap* que já é conhecido. Além disso é notório que o ano mais crítico para que uma *ap* venha a ser anulada é quando esta atinge entre 1 a 2 anos de *tempoap*.

Tabela 3.4 Proporção de *ap* anuladas - não acumulada

		<i>tempoap</i> (anos)						
		0	1	2	3	4	5	6
Ano Criação	2010	8,07%	16,66%	11,89%	9,28%	6,23%	5,91%	5,59%
	2011	5,81%	18,91%	13,72%	10,66%	7,38%	5,67%	
	2012	10,99%	20,98%	13,89%	9,96%	6,95%		
	2013	14,00%	19,87%	12,43%	10,56%			
	2014	11,77%	17,96%	14,16%				
	2015	10,95%	20,88%					
	2016	13,32%						
Média		10,70%	19,21%	13,22%	10,12%	6,85%	5,79%	5,59%

Modelos de Regressão Logística - Estimar a Probabilidade de Anulação de *ap*

O que se sugere para a resolução do problema de retenção de clientes é a aplicação de descontos a *ap* que possam estar prestes a anular. Poderia ser aplicado qualquer outro incentivo para manter as *ap*. No entanto, optou-se por desconto uma vez que se detetou que as condições de *ap* de colaboradores se tornam mais aliciantes para este objectivo.

Deste modo é preciso começar por identificar quais as *ap* que estão prestes a abandonar a seguradora para se aplicar medidas de retenção nas mesmas. Para isso é essencial conhecer a probabilidade de anulação de uma *ap* para se localizar as que estão em “risco”. Servindo este modelo como um alerta para a companhia de Seguros.

No capítulo de “Fidelização” existe uma forte suspeita de que o *tempoap* influencia a probabilidade de a *ap* ser, ou não, anulada (dependendo da fase em que a mesma se encontra). Assim torna-se evidente que a variável *tempoap* é uma variável candidata ao modelo.

O facto anterior levanta um problema: uma vez que na base de dados existem *ap* em vigor e que só se sabe da situação das mesmas até dia 1 de agosto de 2017 está-se perante um caso de dados censurados.

Para contornar a situação, o cálculo da probabilidade de anulação da *ap* foi dividido no seguinte modo:

- Probabilidade de a *ap* anular no ano corrente sabendo que:
 - A *ap* tem menos de 1 ano – modelo P1
 - A *ap* tem entre 1 a 2 anos – modelo P2
 - A *ap* tem entre 2 a 3 anos – modelo P3

Como se considerou que as *ap* se “anulavam” no máximo dia 1 de agosto de 2017 (“congelamento” das *ap* em vigor) o que distingue as mesmas é a variável *vigor* (explicado no capítulo “Variáveis Extra”).

Apesar da divisão de probabilidades, a única distinção é a BD onde cada modelo se baseia. Isto é, a construção de cada modelo é feita sobre a respetiva BD (ou seja, dentro do respetivo intervalo de tempo considerado).

Deste modo a base de dados para a construção de cada modelo foi adaptada ao mesmo. Ou seja, considerou-se apenas as apólices dentro do período de tempo correspondente e adaptou-se a variável resposta, “*vigor*”:

- (BD0) - Para modelo P1:
 - Considerou-se todas as variáveis da BD original e a variável resposta, *vigor*, toma valor “1” para a *ap* anuladas com menos de um ano e “0” caso contrário
- (BD1) - Para modelo P2:
 - Considerou-se apenas as *ap* com *tempoap* superior a 1 e a variável resposta toma o valor “1” se *ap* tiver sido anulada entre 1 a 2 anos e “0” caso contrário
- (BD2) – Para modelo P3:
 - Considerou-se apenas as *ap* com *tempoap* superior a 2 e a variável resposta toma o valor “1” se *ap* tiver sido anulada entre 2 a 3 anos e “0” caso contrário

Resumidamente o cálculo desta probabilidade foi dividida de acordo com os tempos de *ap* conhecidos. Uma vez que, se suspeita, que uma das variáveis mais importantes será a duração da mesma. Pelo que, à partida, *ap* iguais mas com *tempoap* diferentes terão probabilidades de anulação também diferentes.

Assim no conjunto abaixo estão todas as variáveis que são candidatas a incluir no modelo, isto é, todas as variáveis da BD exceto aquelas que teoricamente não fazem sentido nenhum influenciar *vigor*. Conforme dito anteriormente é importante “olhar” para as variáveis antes de “olhar” para os números:

vigor ~ *conddeclarado* + *opcao* + *facc* + *categoria* + *uso* + *valorbm* + *servicoreboque* + *assistviagem* + *colaborador* + *descontagrav* + *valoradescagrav* + *existeacordo* + *sinistros* + *ano1matricula* + *anonascond* + *anocartacond* + *valordesvalorizado* + *cilindrada* + *peso* + *franquia* + *valornovo* + *valorreboque* + *valortomador* + *valorextra* + *numerocd* + *anoinicio*

Variáveis como o género do tomador de seguro ou a zona de circulação não foram consideradas por existirem demasiadas *ap* sem essa mesma informação.

Análise Simples

Na explicação dos resultados só vai ser detalhadamente explicado P1. Para os restantes modelos só vão ser mostradas as conclusões uma vez que a aplicação da metodologia é exatamente a mesma.

A variável resposta é binária toma o valor “1” caso a *ap* esteja anulada (sucesso) e “0” caso a mesma esteja em vigor (insucesso). De qualquer modo esta variável, em termos absolutos, segue a seguinte distribuição (para cada BD respetivamente):

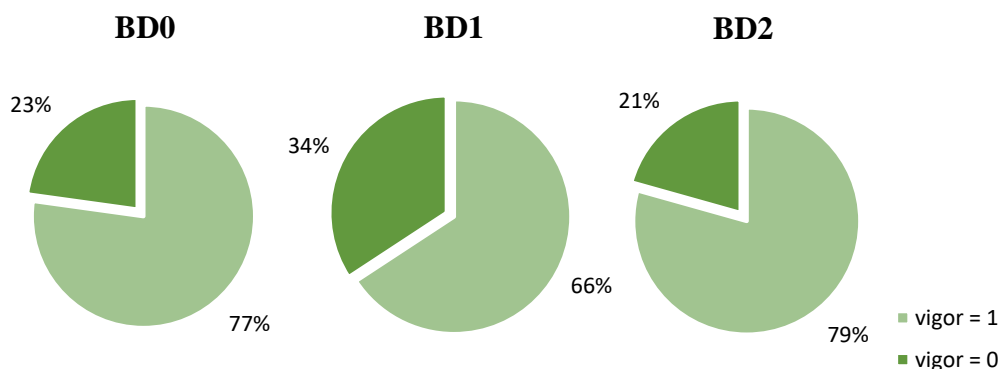


Figura 3.16 Percentagem das *ap* em vigor em cada uma das BD

Partindo-se das variáveis anteriormente identificadas aplicou-se o teste de regressão logística simples. Onde se considerou um alfa de 0.25 para distinguir as possíveis variáveis significativas ou não.

O valor do alfa é bastante superior aos alfas usuais uma vez que se procura um modelo final com mais do que uma variável. Logo a aplicação da regressão simples vai funcionar apenas como uma pré-seleção. Ou seja, a “informação mais útil” que é retirada daqui é a suspeita de quais as variáveis que poderão ser significativas para o modelo.

São retiradas do conjunto de variáveis todas aquelas que não poderão influenciar, a variável resposta. Pelo que o valor do alfa aplicado é bastante elevado para dar uma margem maior na possível significância das variáveis. Repare-se que uma variável pode não ser significativa individualmente mas já o ser na interação com outras, daí que se considere um alfa tão elevado.

Aplicando o modelo de regressão logística simples concluiu-se que:

- Variáveis significativas: *facc*, *existeacordo*, *anonascond*, *anocartacond*, *cilindrada*, *peso*, *franquia*, *valornovo*, *valorextra*, *anoinicio*, *conddeclarado*, *opcao*, *categoria*, *uso*,

valorbm, servicoreboque, assistviagem, colaborador, descontagrav, sinistros, ano1matricula, valorreboque, valortomador, valordesvalorizado e *numerocd*;

- Variáveis não significativas: *valorextra, franquia, valorreboque, valortomador* e *uso*;

Curiosamente a variável *sinistros* influencia negativamente a probabilidade de anular. Este facto não era esperado uma vez (na maioria dos casos) o prémio é agravado, o que não agradará aos tomadores de seguro.

Deste modo foi feita uma análise à proporção de anulação das *ap* por número de sinistros para analisar melhor o peso dos sinistros nas anulações das mesmas.

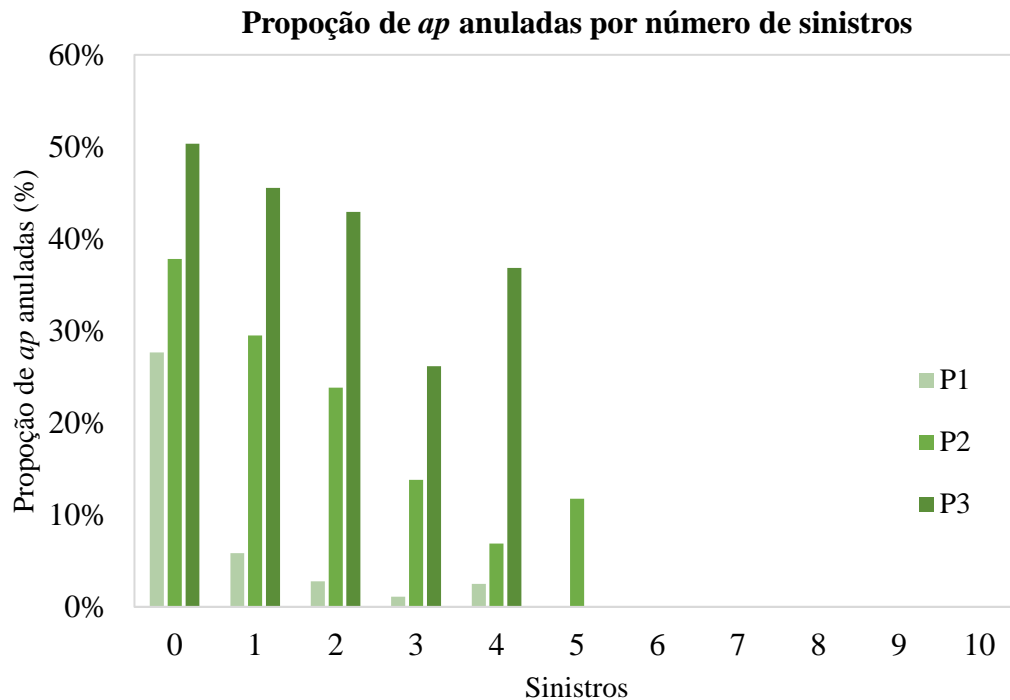


Figura 3.17 Proporção de *ap* anuladas por número de sinistro conhecido

De um modo geral conforme o número de sinistros aumenta a proporção de anulações diminui.

- No caso de 0 sinistros este facto pode ser explicado pela elevada taxa de anulação nas *ap* com menos de 1 ano. Embora seja possível, menos de um ano, é pouco tempo para a ocorrência de sinistros. Logo a maioria das anulações das *ap* desta classe vão ser inseridas nesta situação.
- Relativamente aos outros casos o mesmo pode ser justificado pelo tipo de saneamento aplicado. A BD em estudo não faz distinção entre as anulações de *ap* por vontade da seguradora e por vontade do tomador de seguro. – Anexo 2

No entanto optou-se por deixar que esta variável se mantivesse no modelo uma vez que, na prática, fazia todo o sentido mante-la.

Análise Múltipla

Uma vez filtradas as variáveis, pode-se agora aplicar os métodos de seleção de variáveis: *stepwise*, *backfoward* ou *forward*.

Considere-se que o conjunto de variáveis “filtradas” para esta segunda fase e para P1 foram:

*vigor ~ conddeclarado + opcao + facc + categoria + valorbm + servicoreboque
+ assistviagem + colaborador + descontagrav + existeacordo
+ sinistros + ano1matricula + anonascond + anocartacond
+ valordesvalorizado + cilindrada + peso + valornovo + anoinicio
+ numerocd*

Começou-se por seleccionar as variáveis por 5 métodos diferentes:

- *Full*: incluía todas as variáveis
- *Null*: não tinha nenhuma variável
- *Mod2*: construído a partir do método *setpwise*
- *Mod3*: construído a partir do método *backfoward*
- *Mod4*: construído a partir do método *forward*

Naturalmente os três últimos métodos deram AIC iguais. Optou-se por seleccionar o *mod2* uma vez que era dos que tinha menor AIC. De qualquer forma comparando os valores de AIC entre *mod2* e *null/full* podemos concluir que as variáveis melhoram o modelo uma vez que AIC diminui.

Como a descida de AIC foi tão ligeira levanta a suspeita de que podem continuar a existir variáveis em excesso no modelo, neste caso, em *mod2*.

Na tabela abaixo estão indicados os AIC em cada um dos casos.

Tabela 3.5 AIC de cada processo utilizado para seleção de variáveis				
<i>Full</i>	<i>Null</i>	<i>Mod2</i>	<i>Mod3</i>	<i>Mod4</i>
41.997,41	34.886,41	34.881,36	34.881,36	34.881,36

Uma vez que, aparentemente, *mod2* tinha muitas variáveis procurou-se retirar algumas delas. O método utilizado foi através da comparação do sinal dos $\hat{\beta}_i$ entre a regressão simples e a múltipla mas o modelo em nada melhorou pelo que se manteve o mesmo.

Na interpretação do modelo final (que se encontra na tabela abaixo) vai ser utilizado o conceito de *odds ratio* que não é nada mais que uma razão de chance, ou seja:

- $Odds < 1$ → a probabilidade de o evento ocorrer é menor na primeira característica do que na segunda
- $Odds = 1$ → a característica em estudo é igualmente provável de ocorrer nos dois grupos a comparar
- $Odds > 1$ → a probabilidade é maior na primeira característica do que na segunda

Segue agora os coeficientes estimados assim como a respetiva interpretação:

Tabela 3.6 Ajustamento de *mod2* – P1

Variáveis		$\hat{\beta}_i$	OR	
Conddeclarado	S	0.22533	1.252736	O odds de uma <i>ap</i> com condutor declarado 25% vezes maior quando comparado com o odds de <i>ap</i> sem condutor declarado
Opcao	Máximos			O odds de uma <i>ap</i> de médios diminui cerca de 5% quando comparado com o odds de uma <i>ap</i> de máximos. O mesmo acontece com uma <i>ap</i> de mínimos mas em 12%
	Médios	-0.03328	0.9672677	
	Mínimos	-0.12053	0.8864505	
Facc	S	0.82720	2.286906	Quando o fracionamento de uma <i>ap</i> é semestral aumenta 2 vezes mais quando comparado com um fracionamento anual
Valorbm	Agravamento			O odds de uma <i>ap</i> com desconto diminui 34% quando comparado com o agravamento e aumento 15% mais se a <i>ap</i> não tiver nenhum tipo de desconto ou agravamento em BM
	Desconto	-0.40808	0.6649257	
	Nada	0.13993	1.150193	
Servicoreboque	S	-0.30978	0.7336083	No caso de ter serviço de reboque o odds diminui 27%
Assistviagem	BaseCM			Comparando com BaseCM o odds de BaseLI ou VpLI diminui cerca de 25% enquanto que com o odds de VipCM diminui cerca de 37%
	BaseLI	-0.26825	0.7647166	
	VipCM	-0.45675	0.6333387	
	VipLI	-0.33514	0.7152379	
Colaborador	S	-0.48929	0.6130615	O odds de uma <i>ap</i> de colaborador diminui cerca de 40% relativamente a <i>ap</i> de não colaboradores
Descontagrav	Agravamento			Quando a <i>ap</i> tem desconto a odds aumenta 73% no entanto com desconto sem premio total diminui 60%
	Desconto	0.55289	1.738269	
	Descontosempremiototal	-0.85056	0.4271756	
Existeacordo	1	-0.09915	0.9056069	Existindo acordo a <i>ap</i> diminui a odds em 10%
Sinistros		-1.39847	0.2469745	Com o aumento de sinistros a odds diminui 65%
Ano1matricula		-0.01534	0.9847771	Com a diminuição da idade da viatura a odds diminui 2%
Anonascond		0.00842	1.008456	Com o aumento de cada uma destas variáveis (individualmente) a odds mantem-se praticamente a mesmas sendo que apenas para o valor desvalorizado da viatura tem uma influencia ligeiramente negativa na odds
Peso		0.00016	1.00016	
Valornovo		0.00001	1.00001	
Valordesvalorizado		-0.00002	0.99998	
Numerocd		-0.13395	0.8746338	Com o aumento do <i>numerocd</i> a odds diminui 13%
Anoinicio		0.62677	1.871556	Com o aumento do ano de início da <i>ap</i> a odds aumenta 87%

Qualidade do modelo final

Uma vez selecionadas as variáveis e ajustado o modelo às mesmas testou-se até que ponto o mesmo reflete a realidade. Vão ser usados dois mecanismos: curva ROC, o teste de Hosmer-Lemeshow e comparação de valores estimados e observados.

Começando pela curva ROC este gráfico vai ser bastante importante uma vez que é a área sob a curva que vai avaliar até que ponto o modelo está bem ajustado.

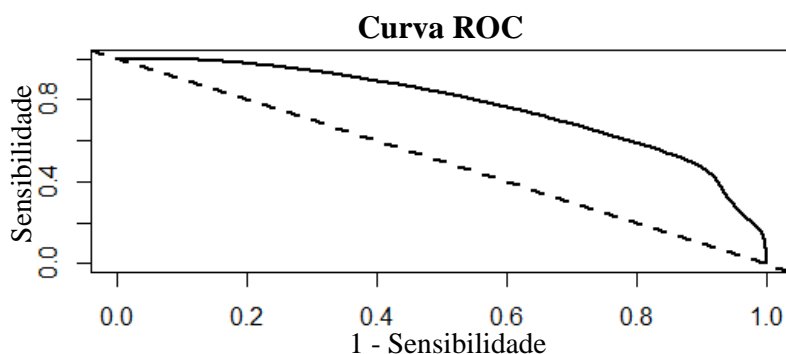


Figura 3.18 Curva ROC para P1

Neste caso a curva ROC de *mod2* tem uma área inferior de 0,77. Logicamente que o ideal é que a área se aproxime de 1. De qualquer modo considera-se que se está perante uma adequação razoável do modelo.

Uma alternativa para testar a qualidade do modelo é através da diferença entre os valores estimados e observados (raciocínio idêntico ao aplicado no teste de Hosmer-Lemeshow).

Assim pode-se concluir que como os pontos a vermelho (realidade) estão muito próximos dos pontos a branco (estimados) está-se perante um modelo relativamente bem ajustado.

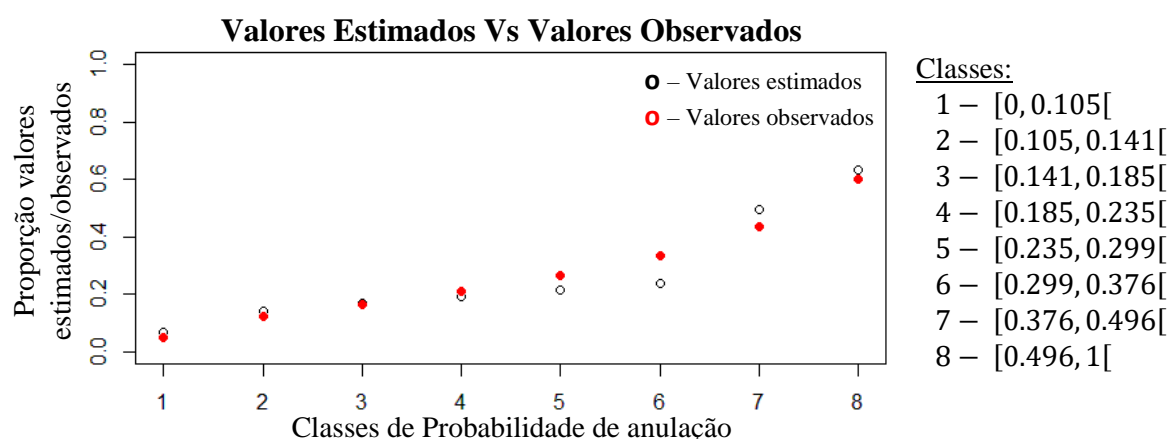


Figura 3.19 Comparação dos valores observados e estimados - P1

No gráfico anterior podemos ver que as classes onde existe maior diferença são as 6 e 7. Estas classes dizem respeito a probabilidades muito próximas de 50% pelo que se torna difícil medir com exatidão a vontade do tomador de seguro na anulação da *ap*. Daí que o modelo também reflita essa indecisão do tomador de seguro.

Restantes Modelos Finais

Assim, para a criação dos restantes modelos (P2 e P3) foi aplicado exatamente o mesmo raciocínio que P1 onde se concluiu que:

- Variáveis significativas:
 - P2: *conddeclarados*, *facc*, *valorbm*, *zona*, *assistviagem*, *existeacordo*, *anoinicio*, *valorbm*, *assistviagem*, *descontagrav*, *ano1matricula*, *anocartacond*, *numeroed* e *sinistros*
 - P3: *facc*, *valorbm*, *descontagrav*, *numeroed*, *anoinicio*, *categoria*, *valortomador* e *sinistros*

No que diz respeito à qualidade dos modelos, para P2 temos que: na curva ROC a sua área inferior é de 0.732 o que considera o modelo razoável.

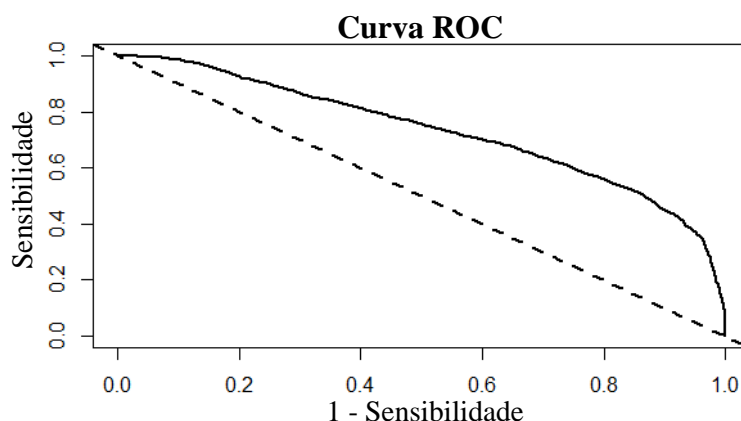


Figura 3.20 Curva ROC para P2

Na comparação dos valores estimados e observados existem umas diferenças maiores. De qualquer modo as maiores diferenças encontram-se nas classes 5 e 6 que correspondem às probabilidades de anular perto dos 50%. Pelo que é normal uma vez que existe uma maior indecisão por parte do tomador de seguro (o que se reflete no modelo).

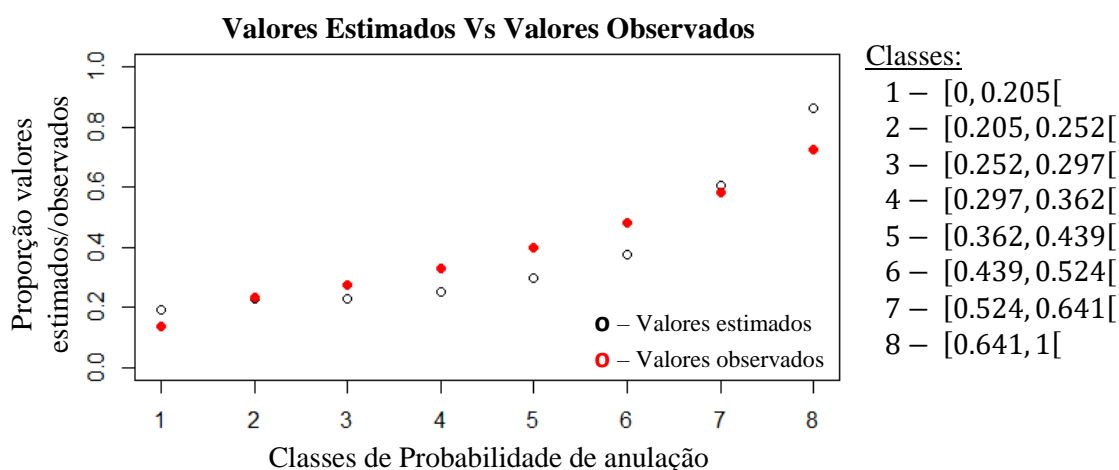


Figura 3.21 Comparação dos valores observados e estimados - P2

O teste de Hosmer e Lemeshow o mesmo também considera o modelo P2 significativo.

No caso de P3 temos que:

- Para a área inferior à curva ROC esta toma um valor de 0.63 (pelo que se considera o menos bom dos três modelos)

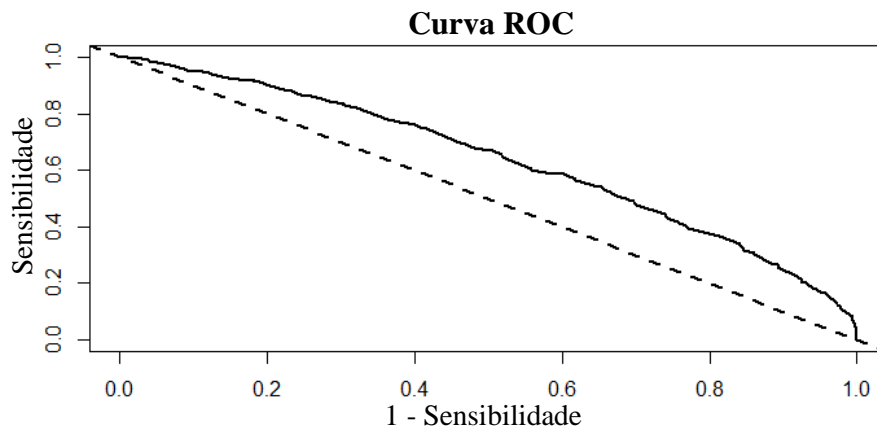


Figura 3.22 Curva ROC para P3

- Na comparação dos valores ajustados e observados, curiosamente, já não se nota a indecisão do tomador de seguro nas classes perto dos 50%

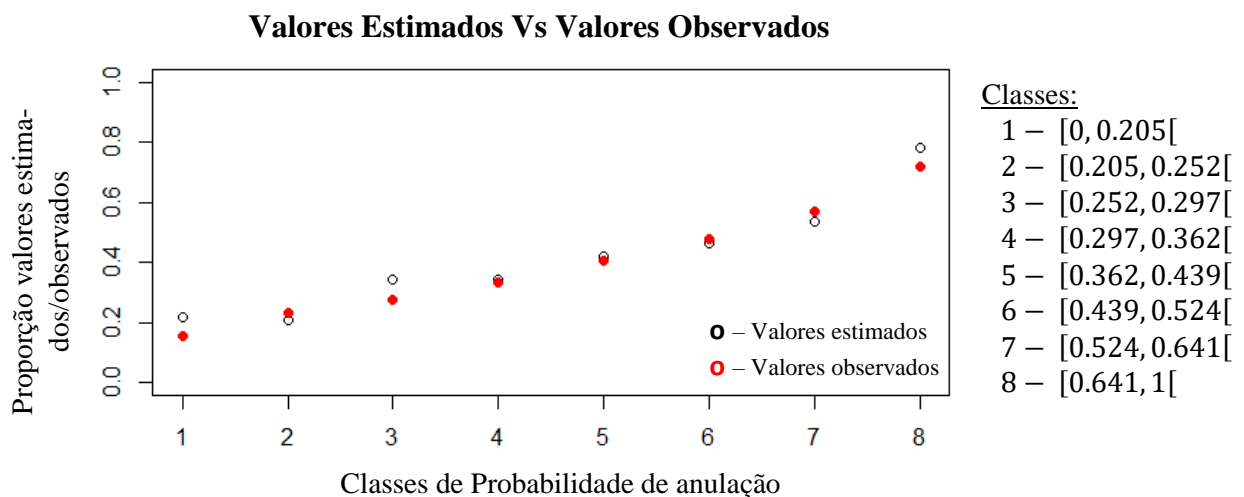


Figura 3.23 Comparação dos valores observados e estimados - P3

Relativamente ao teste de Hosmer e Lemeshow também considera o modelo significativo.

Na comparação dos três modelos – anexo 5 – é natural que existam variáveis significativas diferentes entre eles assim como influencias na anulação contrárias.

No entanto as variáveis *valorbm*, *descontagrav* e *existeacordo* são as únicas nos modelos que se referem a descontos/agravamentos nas *ap* em geral. Mais se nota que as mesmas contribuem positivamente para a anulação quando tomam o valor de descontos.

Este facto parece contraditório aos testes iniciais onde se vê que os descontos de colaboradores ajudam a manter a *ap*. No entanto esta contradição pode ser reflexo do descontentamento do tomador de seguro nos descontos oferecidos.

Repare-se ainda que os descontos de colaboradores ajudam a manter as *ap* (teste feito nos primeiros capítulos) mas os descontos para todas as *ap* (em geral) já não aparentam ser assim tão significativos.

Sinistralidade

Uma vez criado o modelo que alerta para as *ap* que estão em risco de sair surge uma nova questão: “É útil que este cliente, de facto, venha a anular a sua apólice?”

O que se defende é a atribuição de descontos a *ap* prestes a anular, para isso a seguradora é obrigada a investir. Logo é essencial que exista uma excelente selecção de *ap* para não haver investimento em vão.

Para um bom resultado da seguradora, após a aplicação do modelo anterior, é muito importante que o mesmo seja feito apenas sobre as “*ap* boas”. Portanto pressupõe-se que exista um correto saneamento.

De qualquer modo é sempre útil prever quais as *ap* que poderão trazer custos para a companhia de seguros. Assim, conseguindo identificar quais as “*ap* menos boas”, nem é necessariamente mau que estas se anulem (mesmo que não existam motivos para sanear as mesmas).

Neste capítulo, vai ser traçado qual o perfil de uma boa e má *ap*. No entanto esta análise é relativamente superficial uma vez que não existem informações suficientes na BD para devolver um resultado mais consistente.

Em termos de proporção, e considerando todos os dados da BD, temos que:

- Mínimos: cada *ap* tem, em média, 1.16 sinistros (82% do total de apólices)
- Médios: cada *ap* tem, em média, 1.94 sinistros (2% do total de apólices)
- Máximos: cada *ap* tem, em média, 2.19 sinistros (16% do total de apólices)

No entanto apesar de uma determinada opção de *ap* ter maior número de sinistros registados não implica maior ou menor custo. Assim, numa perspectiva de custos temos que:

- Mínimos: custo médio 1.443€
- Médios: custo médio 1.636,30€
- Máximos: custo médio 2.498,80€

Neste caso, de facto, as *ap* com maior número de sinistros têm também custos mais elevados (isto tendo em conta o tipo de opção de cada *ap*).

De qualquer modo, apesar de *ap* de máximos serem mais caras é interessante conhecer quais as coberturas mais dispendiosas.

Tabela 3.7 Coberturas mais cara (custo médio) – sinistralidade

Quebra Isolada de Vidros	536,04€	+, **, ***	
Bagagem	868,24€	++, +++/**	
Responsabilidade Civil Obrigatória	1.558,54€	*	* Opção de mínimos (base)
Atos Vandalismo e Fenómenos Natureza	2.950,74€	**, ***	** Opção de médios (base)
Choque, Roubo ou Incêndio	4.164,24€	***	*** Opção de máximos (base)
Privação de Uso	5.789,98€	++, ***	+ Opção de mínimos (opcional)
Responsabilidade Civil Facultativa	8.104,12€	+, **, ***	++ Opção de médios (opcional)
Acidentes Pessoais Condutor	5.533,63€	+/*, **, ***	+++ Opção de máximos (opcional)

Conforme esperado as coberturas mais caras para a companhia de seguros são as de opções de máximos. No entanto as duas coberturas mais caras podem também ser de mínimos.

Tabela 3.8 Custo médio de cada risco por cobertura - sinistralidade

RESPONSABILIDADE CIVIL OBRIGATÓRIA	
- RCO Danos Corporais	10.002,39€
- RCO Danos Materiais	1.110,26€
RESPONSABILIDADE CIVIL FACULTATIVA	
- RCF Danos Corporais	8.390,10€
- RCF Danos Materiais	7.818,13€
CHOQUE, ROUBO OU INCÊNDIO	
- Choque Colisão e Capotamento	3.962,41€
- Incêndio Raio e Explosão	8.810,34€
- Furto ou Roubo	5.997,04€
- Quebra Isolada de Vidros	414,54 €
QUEBRA ISOLADA DE VIDROS	536,04 €
ATOS VANDALISMO E FENÓMENOS NATUREZA	
- Fenómenos da Natureza	4.626,36€
- Atos de Vandalismo	2.781,39€
ACIDENTES PESSOAIS CONDUTOR	
- Morte ou Invalidez Permanente	7.509,15€
- Despesas de Tratamento	8.566,03€
PRIVAÇÃO DE USO	5.789,97€
BAGAGEM	868,24 €
CUSTOS DE GESTÃO	1.498,07€
CUSTOS DE GESTÃO DC	7.304,27€

Analisando os custos de cada cobertura é visível que os custos para danos corporais são significativamente superiores aos de danos materiais (o que é normal). Repare-se ainda que a cobertura mais cara corresponde apenas a riscos de danos corporais.

De qualquer modo o risco mais dispendioso: incêndio, raio ou explosão (cobertura esta que pertence a *ap* de máximos/médios), não pertence à cobertura mais cara.

Conforme explicado no primeiro capítulo existem dois tipos de regularização de sinistros que se distribuem do seguinte modo:

- Não IDS (pode atribuir qualquer percentagem de responsabilidade do sinistro à viatura segura – depende apenas do sinistro em si)
- IDS (sinistros ocorridos dentro de condições específicas):

- Devedor (apenas sinistros onde se atribui 100% de culpa à viatura segura)
- Misto (apenas sinistros onde se atribui qualquer percentagem de culpa à viatura segura exceto 0% e 100%)
- Credor (apenas sinistros onde se atribui apenas 0% de culpa à viatura segura)

Deste modo o tipo de sinistros participados à companhia de seguros da BD em estudo distribui-se da seguinte maneira:

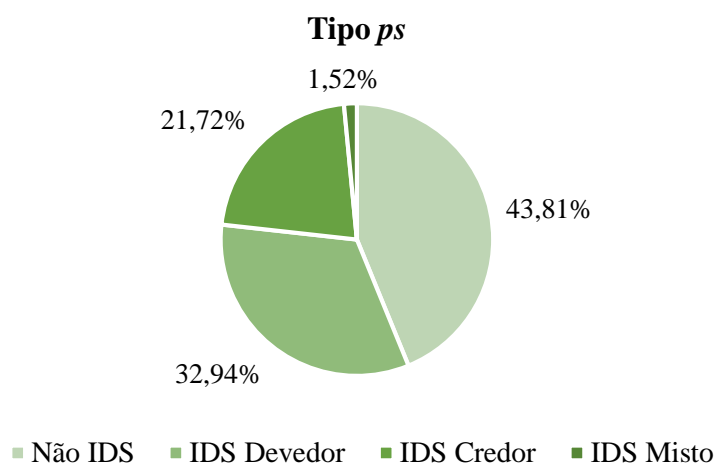


Figura 3.24 Percentagem de tipo de *ps* participados/reclamados à seguradora

No entanto uma maior percentagem não implica maior custo. Portanto analisando o custo por tipo de *ps* temos que:

Tabela 3.9 Custo médio e percentagem por *ps*

Custo por tipo de <i>ps</i>		% de tipo de <i>ps</i> participado	
IDS Credor	176,68€	IDS Misto	1,52%
IDS Misto	564,39€	IDS Credor	21,72%
IDS Devedor	1.507,77€	IDS Devedor	32,94%
Não IDS	3.098,66€	Não IDS	43,81%

Note-se que a tabela anterior está por ordem crescente (quer no custo quer na percentagem), e como se pode ver a distribuição do tipo de sinistros não é igual à da percentagem. Sendo que sinistros do tipo não IDS são também os que têm maior custo e maior frequência de participação.

Repare-se que nos sinistros do tipo IDS Credor o valor médio não é nulo, apesar da viatura segura não ter culpa. Este facto pode parecer estranho uma vez que, com inexistência de culpa, deveria haver inexistência de custos. Neste tipo de *ps* além de ter custos de gestão, o custo do sinistro em si é depois regularizado na câmara de compensação que funciona à base de médias. Portanto raramente a companhia de seguros recebe o valor real do custo do sinistro, daí que este custo não seja nulo.

Os sinistros do tipo Não IDS este valor está bastante inflacionado devido aos sinistros com danos corporais que (como foi visto antes) têm valores bem superiores aos sinistros de danos materiais. Seria ainda interessante analisar os custos deste tipo de *ps* por percentagem de culpa. No entanto esta informação não existe na BD pelo que não foi possível fazê-la.

Relacionando os dois aspetos anteriores (tipo de sinistro e opção de *ap*) qualquer que seja o tipo de *ps* a maioria das *ap* tem mínimos, de seguida temos *ap* de médios e, por fim, de máximos. Sendo que, se nota um aumento de percentagem nas *ap* de máximos no tipo de *ps* não IDS comparativamente com os restantes tipos de *ps*.

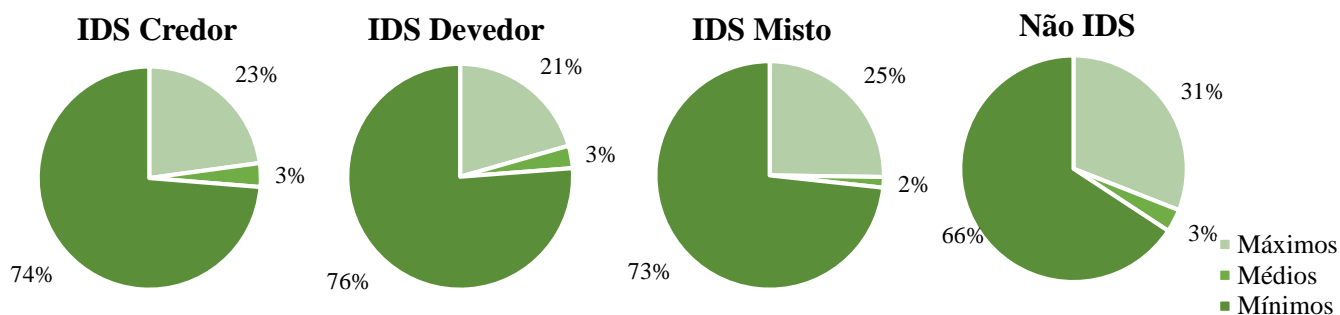


Figura 3.25 Percentagem de opção de *ap* por tipo de *ps*

A informação anterior era esperada uma vez que os resultados obtidos seguem a distribuição da quantidade de *ap* que existe na BD. Ou seja, como existem mais *ap* de mínimos é normal que existam mais sinistros ocorridos com *ap* de mínimos (analogamente para as restantes opções).

Devido ao facto anterior foi feita uma análise mas ao contrário, ou seja, a distribuição do tipo de *ps* para cada opção de *ap*.

De qualquer modo a distribuição por tipo *ps* é toda ela semelhante entre as várias opções de *ap*. É natural que, entre *ps* IDS, a maioria dos sinistros participados seja do tipo IDS credor uma vez que é um sinistro que não implica agravamentos na *ap* logo o tomador de seguro não evita participar à seguradora.

A percentagem de *ps* do tipo não IDS em *ap* de máximos têm uma percentagem superior aos restantes tipos de *ap*. Isto acontece uma vez que os sinistros de danos próprios são sempre regularizados por esta via. Note-se que as *ap* de mínimos são os que têm menor percentagem do tipo não IDS uma vez que esta modalidade de *ap* não tem, de todo, coberturas de danos próprios.

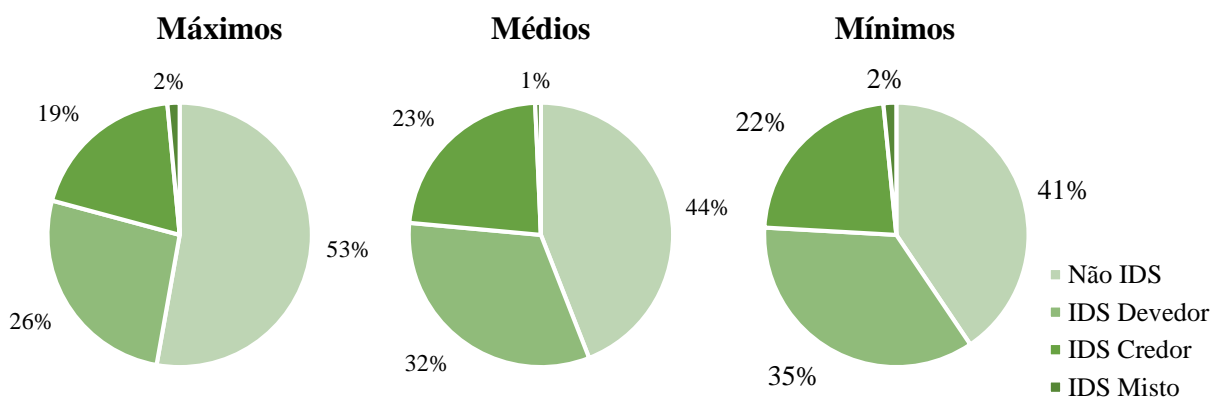


Figura 3.26 Percentagem de opção de *ap* por tipo de *ps*

Embora não tenha assim tanta importância para o perfil de *ap* “boas/más”, a título de curiosidade, analisou-se qual a altura de maior sinistralidade. Deste modo concluiu-se que a participação de sinistros atinge os seus picos durante os períodos escolares, ou seja:

- 1º período: setembro a dezembro;
- 2º período: janeiro a abril (início);
- 3º período: abril (final) a junho;

Como é visível os valores mínimos são: abril, agosto e início do ano (janeiro/fevereiro). Isto acontece pelo facto de, assumindo que a maioria dos *ps* se trata de um “pequeno” sinistro (e que a viatura “funciona”), as pessoas esperaram para regressar das férias e tratar destes “problemas”. Daí que os picos ocorram durante a conhecida “época baixa”.

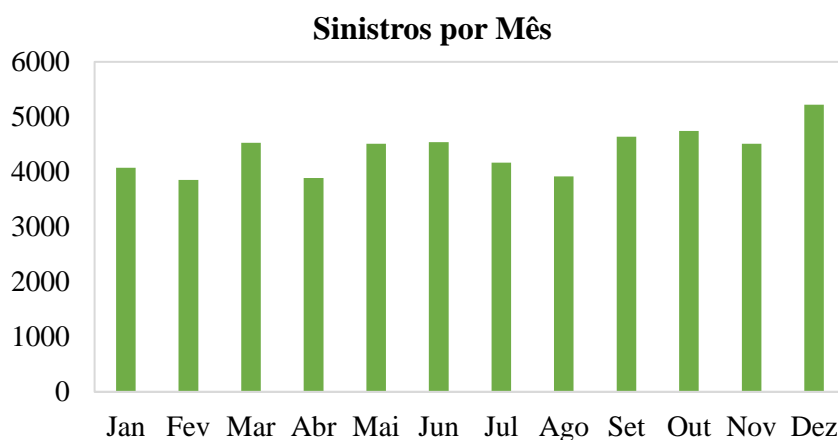


Figura 3.27 Ocorrência de sinistros mensal

Resumindo, pode-se concluir então que as *ap* que acarretam mais custos à companhia de seguros são *ap* de máximos. Essencialmente se forem *ap* que tenham seguras as seguintes coberturas:

- Privação de uso;
- Responsabilidade civil facultativa;
- Choque, Roubo ou Incêndio;

Apesar de a cobertura de acidentes pessoais ser a mais dispendiosa esta existe em qualquer uma das opções de *ap* (o que varia é o limite de capital) por isso não faz sentido estar a distinguir as *ap* (no que diz respeito aos custos) por esta cobertura.

Aplicações

A maior utilidade dos modelos anteriores é possibilitar à companhia de seguros identificar as *ap* em risco e atuar de modo a retê-las na mesma.

Um dos métodos que se sugere para “segurar o cliente” seria a aplicação de descontos extra. Conforme visto no início, existem diferenças significativas entre *ap* com descontos “especiais” e sem eles.

Embora para o cliente parecesse um desconto atribuído “por sorte”, a seleção das *ap* que iriam usufruir do mesmo iria depender da probabilidade de anulação que fosse estimada.

Para atribuição de descontos estabeleceu-se a seguradora iria atribuir os mesmos para probabilidades de anulação acima de 70% - ponto de corte. Quanto menor for a esta probabilidade menos dinheiro a companhia de seguros vai receber. Pelo que este valor vai depender também de quanto se está disponível a investir.

O mesmo raciocínio pode ser aplicado mas exatamente no sentido contrário, ou seja, agravar prémios de *ap* se as mesmas têm uma probabilidade de anular muito baixa. Esta probabilidade sendo baixa significa que se trata de uma *ap* mais “segura”. Assim a companhia de seguros tem uma margem maior para “apertar” com o cliente, visto que este “suporta”.

Este método é também uma maneira, não só de a seguradora ganhar dinheiro, mas também de compensar os descontos aplicados nas outras *ap* em risco. Considere-se então, para os agravamentos, um ponto de corte de 2%.

Relativamente aos pontos de corte considerados, o ideal é:

- Nos descontos: se maximize os VP, diminuindo os FN assim como os FP. Uma vez que se está a atribuir descontos a *ap* que não iam anular (FP) em vez de atribuir a *ap* que, de facto, o iam fazer (FN);
- Nos agravamentos: análogo ao caso anterior mas maximizando os VN e diminuir FN e FP;

Deste modo chegou-se aos seguintes resultados (para P1):

- Descontos: 21,5% correspondia a *ap* que iriam anular mas que não foram “abrangidas” pelo desconto, FN, e 0,1% correspondia a *ap* que iria usufruir do desconto mas desnecessariamente, FP;
- Agravamentos: 0,1% corresponderia a *ap* que foram dadas como não anuladas e anularam, FN, e 70% correspondia a *ap* que foram dadas para anular e não anularam, FP;

Tabela 3.10 FN e FP nos modelos P2 e P3. Considerando um ponto de corte de 70% para descontos e 2% para agravamentos

	Descontos			Agravamentos		
	P1	P2	P3	P1	P2	P3
FN (ajust = 0 e obs =1)	21,5%	29%	44%	0,1%	0%	0%
FP (ajust = 1 e obs =0)	0,1%	0,5%	0,2%	70%	65%	48%
Sensibilidade	0.054	0,153	0,06	1	1	1
Especificidade	0,999	0,992	0,997	0	0,003	0

Para o ponto de corte o que se procurou foi: nos agravamentos minimizar os FN para não “prejudicar” *ap* “boas” e nos descontos minimizar os FP para não levar a um investimento em vão por parte da companhia de seguros.

Depois de identificadas as *ap* a atuar (de acordo com o ponto de corte) o que se pretende é eliminar aquelas que têm características de baixa/alta sinistralidade (dependendo se a aplicação seria de descontos ou agravamento).

Resumidamente a estrutura de ação seria:

1. Correr os modelos P1, P2 e P3 para as respectivas *ap*, ou seja, *ap* com menos de 1 ano, *ap* com 1 a 2 anos e *ap* com 2 a 3 anos respetivamente;
2. Identificar um ponto de corte e selecionar as *ap* acima/abaixo desse valor para aplicar descontos/agravamentos;
3. No grupo de *ap* para, eventualmente aplicar descontos, filtrar aquelas que têm características que com baixa sinistralidade – essas *ap* que se filtrou são as *ap* a atuar;
4. No grupo de *ap* para, possivelmente aplicar agravamentos, fazer o contrário do caso anterior. Ou seja, filtrar as *ap* com características de alta sinistralidade – essas mesmas *ap* são as que se vão aplicar os agravamentos;

Relativamente ao valor do desconto o objetivo é que seja apenas o suficiente para “segurar” o cliente, tendo em conta a previsão de quanto a *ap* em causa pode dar à empresa. No caso dos agravamentos este deve ser aplicado de modo a compensar o máximo possível o valor que se descontou nas *ap* anteriores sem por em causa a anulação das *ap* escolhidas.

Seria importante calcular quanto é que um 1% de desconto/agravamento influencia na probabilidade de a *ap* anular, e a partir de toda esta informação, deduzir o valor ideal do mesmo.

4. CONCLUSÃO

Em conclusão, detetou-se uma elevada falta de fidelização devido às grandes taxas de anulação. No entanto nota-se uma diferença significativa entre apólices de colaboradores e de não colaboradores. O que seria esperado uma vez que as apólices de colaboradores têm uma tarifa diferente tornando-as mais atrativas.

Perante esta situação sugere-se a aplicação de descontos a apólices que estejam a anular e agravamentos caso contrário. Para este efeito são construídos três modelos que quantificam a probabilidade de uma apólice anular em três casos:

- Sabendo que a apólice tem menos de um ano – P1
- A apólice tem entre 1 a 2 anos – P2
- A apólice tem entre 2 a 3 anos – P3

Esta divisão foi feita uma vez que se identificou uma certa estabilização da taxa de anulação com o passar dos anos. Portanto deduz-se que duas apólices com características iguais mas com durações diferentes terão probabilidades de anular também diferentes.

Uma vez que o trabalho se baseia numa amostra onde existem apólices em vigor está-se perante a existência de dados censurados. A criação dos três modelos, distinguidos pela duração já conhecida de cada apólice, deve-se também a este facto.

O último modelo criado vai apenas até apólices entre 2 a 3 anos porque além de o tempo médio ser, aproximadamente, os 2 anos a tal estabilização da taxa de anulação começa também perto desse ano.

Relativamente aos modelos, o que tem melhor qualidade é P1 no entanto considera-se que todos eles têm uma aproximação razoável à realidade.

Através destes modelos já se consegue identificar quais as apólices com maior/menor risco de anular o que facilita aplicação dos descontos/agravamentos indicados no início.

No entanto podem existir apólices que até sejam convenientes que se anulem, uma vez que se prevê que possam trazer muitos custos à seguradora. Assim fez-se também uma análise à sinistralidade. Esta análise foi feita com base nos custos de sinistros e não, necessariamente, na quantidade. Uma vez que, nem sempre muitos sinistros levam a custos mais elevados.

Com base nos modelos e nas características de alta/baixa sinistralidade já se faz uma melhor seleção de apólices para retenção de clientes.

No que diz respeito às limitações e comentários a este trabalho temos que:

- A amostra tem duração de apenas 7 anos o que para o estudo em si é pouco tempo;
- Poucas variáveis a nível de sinistralidade para fazer uma análise mais consistente sobre o assunto;
- Não existe distinção entre as apólices que anularam por vontade própria e as que anularam por vontade da companhia de seguros. Este estudo recai apenas sobre o primeiro caso, portanto seria essencial que a análise recai-se apenas sobre estes – este facto justifica parte da variabilidade existente;
- os modelos construídos – P1, P2 e P3 - além de se complementarem são também adaptáveis com o aumento/diminuição de variáveis;

- Pressupõe-se que é aplicado um bom saneamento pelo que o modelo é apenas aplicado sobre apólices “boas”;

Para um aperfeiçoamento deste trabalho sugere-se:

- Analisar o peso das *ap* que são criadas e automaticamente anuladas para inclusão ou não na BD e aplicação de medidas contra esses casos se necessário;
- Criação de um modelo para identificar a probabilidade de ocorrer um sinistro (análogo ao modelo criado para a probabilidade de a apólice anular). Com a interação dos dois modelos as apólices escolhidas para aplicação dos descontos/agravamentos seria mais justa;
- Uma vez que se trata de um estudo social seria muito interessante incluir uma variável que medisse a interação entre a companhia de seguros e o tomador de seguro. Ou seja, uma pessoa que, recorrentemente, entrasse em contacto com a seguradora teria um valor diferente de outra que raramente tivesse contacto com a mesma. No entanto é complicado fazer uma medição justa desta variável uma vez que se trata de uma matéria mais abstrata;
- Avaliar se não existem mais variáveis significativas como o histórico de sinistralidade, estado civil, agregado familiar...;
- Avaliação sobre o valor de desconto/agravamento a aplicar – medir até que ponto 1% influencia;

O intuito desta tese é devolver uma boa estrutura que adaptando ao caso real de cada companhia de seguros se consiga aumentar a retenção de clientes da mesma.

Uma vez que se sugere a atribuição de descontos/agravamentos por conveniência da companhia de seguros levanta-se uma questão: até que ponto esta prática é ética? Até que ponto é aplicado o direito de igualdade entre os tomadores de seguro?

Embora aparentemente exista uma distinção entre apólices essa distinção é feita sobre todas elas. Ou seja, os tais descontos/agravamentos podem ser atribuídos a qualquer cliente pelo que não existe discriminação entre tomadores de seguros.

5. BREVE GLOSSÁRIO DE SEGUROS

Acidente de viação: sinistro ocorrido em consequência exclusiva da circulação rodoviária, independentemente de o veículo seguro estar ou não em movimento, quando condutor se encontre dentro dele, a entrar ou a sair ou a participar de forma ativa, no decurso de uma viagem, em trabalhos de pequena reparação ou desempanagem desse veículo.

Atos de vandalismo: danos causados ao veículo seguro por pessoa ou pessoas agindo maliciosamente.

Apólice de seguro: conjunto de condições devidamente identificado e na qual é formalizado o contrato de seguro celebrado.

Beneficiário de um contrato de seguro: Pessoa ou entidade com direito às prestações previstas no contrato de seguro.

Bonificação ou bônus: Diminuição do prémio na renovação do contrato de seguro, nas situações fixadas na apólice (por exemplo, não terem ocorrido sinistros).

Capital garantido: Cláusula contratual nos termos da qual o segurador se obriga a reembolsar o investimento no prazo acordado em montante não inferior à totalidade do capital inicialmente investido.

Capital Seguro: montante máximo indicado nas condições da apólice, a pagar pelo Segurador a título de indemnização; no caso das coberturas choque, colisão ou capotamento, incêndio, raio ou explosão, furto ou roubo, vale como limite indemnizatório tanto para a perda total como para a perda parcial e é atualizado periodicamente.

Carta Verde: Documento comprovativo da existência do seguro obrigatório de responsabilidade civil de automóvel, em termos válidos e eficazes, também designado por certificado internacional de seguro. A carta verde, enquanto prova da existência de seguro, é válida em todos os países nela mencionados.

Carteira de Seguros: Conjunto de contratos de seguro em relação aos quais o mediador de seguros exerce a atividade de mediação e que lhe criam direitos e deveres para com seguradores e tomadores de seguros.

Choque: embate do veículo contra qualquer corpo fixo, ou embate sofrido por aquele quando imobilizado.

Cobertura: Conjunto de situações cuja verificação determina a prestação do segurador ao abrigo do contrato.

Colisão: embate entre o veículo e qualquer outro corpo em movimento.

Condições especiais: Disposições que completam ou especificam as condições gerais, sendo de aplicação generalizada a determinados contratos do mesmo tipo.

Condições gerais: Cláusulas que são acrescentadas às condições gerais/especiais de um contrato, para o adaptar a um caso particular, precisando nomeadamente o risco coberto, a duração e o início do contrato, o capital seguro, o prémio, o tomador do seguro, o segurado e o beneficiário.

Condutor: o legítimo condutor do veículo seguro no momento em que este intervenha num acidente de viação.

Congénere: Companhia de seguros que regulariza danos da viatura terceira em caso de sinistro. Qualquer empresa de seguros que não a do próprio tomador de seguro.

Contrato de seguro: Contrato através do qual o segurador assume a cobertura de determinados riscos, comprometendo-se a satisfazer as indemnizações ou a pagar o capital seguro em caso de ocorrência do sinistro, nos termos acordados. Em contrapartida, o tomador do seguro obriga-se a pagar o prémio correspondente.

Dano corporal: prejuízo resultante de lesão da saúde física ou mental. Dano relativo à vida, à saúde ou à integridade física de uma pessoa.

Dano Material: prejuízo resultante de lesão de coisa móvel, imóvel ou animal. Prejuízo causado a coisas, bens materiais, créditos e quaisquer outros direitos patrimoniais.

Declaração Amigável de Acidente Automóvel (DAAA): impresso a preencher em caso de acidente automóvel. Destina-se a recolher certas informações indispensáveis à regularização do sinistro pelos seguradores e a fazer a participação do acidente. Este impresso, sempre que possível, deve ser preenchido imediatamente no próprio local do acidente e assinado por ambas as partes.

Despesas de tratamento: Despesas relativas a honorários médicos e internamento hospitalar, assim como assistência medicamentosa e de enfermagem, que forem necessários em consequência de acidente.

Empresa de Seguros: Entidade legalmente autorizada a exercer a atividade seguradora e que subscreve, com o tomador de seguro, o contrato de seguro.

Estado membro: os Estados subscritores do Acordo sobre o Espaço Económico Europeu, de 2 de Maio de 1992.

Existências: Quantidade de mercadorias, de bens ou de objetos diversos cujo valor está seguro total ou parcialmente.

Explosão: ação súbita e violenta da pressão ou depressão de gás ou de vapor.

Fenómenos da Natureza: acontecimentos de carácter excecional que provoquem danos no veículo seguro e que decorram de:

- Tempestades com ventos de velocidade superior a 100 km/hora, em contínuo ou em rajada, comprovada por documento emitido pela estação meteorológica mais próxima, que provoquem a queda de quaisquer objetos tais como árvores, telhas, chaminés, muros ou similares
- Chuvas torrenciais e trombas-d'água com precipitação de intensidade superior a 10 mm em 10 minutos, comprovada pela estação meteorológica mais próxima, que provoquem inundações, enxurradas, rebentamento de adutores, coletores, diques, barragens e similares;
- Tremores de terra, erupções vulcânicas e outros fenómenos sísmicos e geológicos, nomeadamente aluimentos, deslizamentos e afundamentos de terrenos.

Fracionamento do prémio: Divisão contratual de um prémio anual em frações, pagas periodicamente.

Franquia: valor da regularização do sinistro nos termos do contrato de seguro que não fica a cargo da companhia de seguros, ou seja, valor correspondente à parte do prejuízo que fica a cargo do tomador de seguro. Ou seja, dano ou parte do dano que fica convencionalmente a cargo do segurado.

Furto: apropriação ilegítima do veículo seguro, incluindo o furto de uso (entendido como a utilização do veículo seguro contra a vontade do Segurado).

Garantia: Âmbito do compromisso, pela empresa de seguros, na cobertura de um risco.

Incêndio: combustão accidental, com desenvolvimento de chamas, estranha a uma fonte normal de fogo, ainda que nesta possa ter origem, e que se pode propagar pelos seus próprios meios.

Indemnização: Valor pago por uma empresa de seguros para reparar ou ressarcir um dano resultante de um sinistro.

Mercadoria: é toda a coisa apreciável economicamente, ou seja, capaz de ter o seu valor convertido em dinheiro (sentido amplo). Para o seguro de Transportes é toda a coisa, objeto do comércio, que é transportada.

Participação: Documento pelo qual o segurado comunica à empresa de seguros a ocorrência de um sinistro, indicando as suas causas, a data, o local, os prejuízos prováveis, etc.

Perda parcial: danificação parcial do veículo seguro que permite a sua reparação, com peças novas, até ao limite do capital seguro considerado para efeitos de perda total. O montante da indemnização paga em caso de Perda Parcial, será abatido ao capital/valor seguro, ficando este reduzido daquele valor desde a data do sinistro até ao vencimento do contrato, para efeitos de nova perda parcial. O Tomador de seguro pode repor o capital/valor seguro através do pagamento de um prémio suplementar correspondente ao capital/valor seguro reposto e no período de tempo não decorrido até ao vencimento anual do contrato.

Perda total: considera-se perda total do veículo seguro:

- a sua destruição total;
- a sua destruição parcial:
 - Quando a reparação não seja materialmente possível ou,
 - Quando a reparação não seja tecnicamente aconselhável ou,
 - Desde que o valor da reparação exceda 80% do Capital seguro atualizado, salvo convenção em contrário expressa nas condições particulares

Em resumo considera-se uma perda total como toda a situação em que o bem seguro sofre danos cujo custo de reparação após o sinistro, acrescido do valor do salvado, ultrapassa o valor venal antes do sinistro.

Prémio: O prémio bruto acrescido das cargas fiscais e parafiscais, e que corresponde ao preço pago pelo tomador de seguro à empresa de seguros pela contratação do seguro.

Provisão para sinistros: Provisão que corresponde ao custo total estimado que a empresa de seguros suportará para regularizar todos os sinistros que tenham ocorrido até ao final do exercício, quer tenham sido comunicados ou não, após dedução dos montantes já pagos respeitantes a esses sinistros.

Raio: descarga atmosférica ocorrida entre a nuvem e o solo, consistindo num ou mais impulsos de corrente que conferem ao fenómeno uma luminosidade característica (raio), e que provoque deformações mecânicas permanentes no veículo seguro.

Ramo (de seguro): Conjunto de operações ou atividades relativas a contratos de seguro da mesma natureza. Por exemplo, ramo vida e ramo não vida.

Reclamação: Pedido de indemnização, apresentado amigavelmente ou por via judiciária, por um terceiro lesado ou pelos seus titulares de direito, à empresa de seguros que cobre o responsável pelo dano.

Regularização/liquidação de sinistro: Acordo sobre o montante definitivo da indemnização, após um sinistro, entre a empresa de seguros e o beneficiário.

Responsabilidade de sinistro: a obrigação de cada um em responder perante outrem por ofensas causadas.

Risco: Eventualidade de ocorrência de um evento aleatório suscetível de afetar o património do segurado. Conjunto de eventualidades consideradas pelas empresas de seguros como fazendo parte de uma mesma categoria, por exemplo, risco de acidente, risco de incêndio, risco de transporte, etc.

Roubo: apropriação ilegítima do veículo seguro mediante o uso de violência ou ameaça para a integridade física ou para a vida do Segurado.

Salvado: Bem (no ramo automóvel parte do veículo) que conserva um certo valor após a ocorrência de um sinistro.

Segurado: a pessoa ou entidade titular do interesse seguro. É sobre quem recai o risco segurado. Pessoa singular ou coletiva no interesse da qual o contrato de seguro é celebrado, ou a pessoa (pessoa segura) cuja vida, saúde ou integridade física se segura.

Segurador: entidade legalmente autorizada para a exploração do seguro obrigatório de responsabilidade civil automóvel, que subscreve os vários tipos de contrato de seguro. (o mesmo que Empresa de seguros)

Segurar: para uma empresa de seguros, comprometer-se por um contrato de seguro a fornecer as prestações previstas em caso de ocorrência de um risco. Transferência de um risco para uma empresa de seguros.

Seguro contra terceiros: (o mesmo que Seguro de responsabilidade civil automóvel)

Seguro de danos próprios: Seguro que garante a reparação ou a substituição de um veículo terrestre após choque, colisão, capotamento, incêndio, raio ou explosão e furto ou roubo.

Seguro de responsabilidade civil automóvel: Seguro de responsabilidade civil que cobre os danos causados a terceiros por veículos terrestres a motor e seus reboques.

Sinistro: a verificação, total ou parcial, do evento que desencadeia o acionamento da cobertura do risco prevista no contrato, considerando-se como um único sinistro o evento ou série de eventos resultantes de uma mesma causa. Evento ou série de eventos resultantes de uma mesma causa suscetível de fazer funcionar as garantias de um ou mais contratos de seguro.

Sociedade de seguros: (o mesmo que Empresa de seguros)

Tarifa: Designação dada ao quadro de prémios ou de taxas de prémio a aplicar aos riscos a segurar e a o conjunto de condições de subscrição de um dado ramo.

Terceiro: aquele que, em consequência de um sinistro coberto pelo contrato de seguro, sofra um dano suscetível de, nos termos da lei civil e da apólice, ser reparado ou indemnizado. Vítima de um sinistro que não é parte no contrato de seguro mas que, por força deste, assume o direito de ser indemnizado.

Titular do contrato: (o mesmo que Tomador de seguro)

Tomador de seguro: Pessoa singular ou coletiva que, por sua conta ou por conta de uma ou várias pessoas, celebra o contrato de seguro com a empresa de seguros, sendo responsável pelo pagamento do prêmio.

Valor venal: Valor comercial de um bem (no ramo automóvel de uma viatura) imediatamente antes da ocorrência do sinistro.

Veículo seguro: veículo automóvel ligeiro ou misto, de matrícula portuguesa, identificado nas condições da apólice, que não efetue serviço público, aluguer (ainda que de curta duração), transporte de animais, ou serviço de ambulância, táxi, instrução ou carreta funerária.

Vigência: Período de validade de uma apólice, pelo qual a empresa de seguros recebeu o prêmio.

6. BIBLIOGRAFIA

- Abraham, B. & Ledolter, J. (2006) *Introduction to Regression Modeling*. Thomson Brooks/Cole
- Almeida, M.D. (1971). *O contrato de Seguro no Direito Portugês e Comparado*. Lisboa: Livraria Sá da Costa.
- Chichorro, M. M. (2010). *O Contrato de Seguro Obrigatório de Responsabilidade Civil Automóvel* (1 ed.). (W. K. Portugal, Ed.) Coimbra, Antanol, Portugal: Coimbra Editora.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. Wiley.
- Kleinbaum, D., Kupper, L., & Muller, K. (1988). *Applied Regression Analysis and Other Multivariate Methods*. USA: Duxbury Press.
- Insurance Europe. (2010). *How Insurance Works*. Bruxelas: Insurance Europe.
- Lopes, M. C. (1987). *Seguro Obrigatório de Responsabilidade Civil Automóvel*. Impresa Nacional - Casa da Moeda.

7. ANEXOS

Anexo 1 – Tabela de Desvalorização da Viatura Segura

Tabela 7.1 Tabela de desvalorização da viatura segura

Idade Viatura	% Desvalorização
1	24
2	35
3	44
4	52
5	59
6	65
7	70
8	74
9	77
10+	79

Anexo 2 – Tabela de Saneamento considerada

Tabela 7.2 Regras de saneamento

3 sinistros QIV em menos de 1 ano	<ul style="list-style-type: none"> Se tem Opção Máximos -> Alterar para Opção Mínimos SEM QIV Se tem Opção Médios com QIV -> Alterar para Opção Mínimos SEM QIV Se tem Opção Mínimos -> Alterar para Opção Mínimos SEM QIV
3 sinistros Fenómenos Natureza e/ou Actos Vandalismo em menos de 1 ano	<ul style="list-style-type: none"> Se tem Opção Máximos -> Alterar para Opção Mínimos SEM QIV Se tem Opção Médios -> Alterar para Opção Mínimos COM OU SEM QIV (Consoante já tenha ou não)
2 sinistros de RC e/ou CCC em menos de 1 ano	Agravar BM
+ 2 sinistros de RC e/ou CCC em menos de 2 anos	Anular apólice
3 sinistros de RC e/ou CCC em menos de 3 ou 4 anos	Anular apólice

Anexo 3 – Franquias por coberturas

Tabela 7.3 Franquia aplicada a cada cobertura contratada

Cobertura	Franquia
Responsabilidade Civil Obrigatória	S/ franquia (*)
Responsabilidade Civil Facultativa	S/ franquia (*)
Assistência em viagem	25 €
despesas médicas, cirúrgicas,	S/ franquia
medicamentos e hospitalares	S/ franquia
Proteção Jurídica	
Acidentes Pessoais Condutor	2%, 10%, 20% (**)
Choque, Roubo e Incêndio	S/ franquia
Choque, Colisão ou Capotamento	2%, 10%, 20% (**)
Furto ou Roubo	S/ franquia
Incêndio, Raio ou Explosão	
Quebra Isolada de Vidros	2%, 10%, 20% (**)
Atos de Vandalismo e Fenómenos da Natureza	2%, 10%, 20% (**)
Atos de Vandalismo	
Fenómenos da Natureza	
Bagagem	S/ franquia
Privação de Uso	3 dias caso de roubo

Anexo 4 – Coberturas e Modalidades de *ap*

Apesar do seguro de Responsabilidade Civil ser o único obrigatório existem outras coberturas facultativas que se podem adquirir de modo a complementar a apólice de seguro e garantir uma maior segurança relativamente aos riscos a que o segurado se pretende “proteger”:

- **Assistência de viagem:** esta cobertura inclui um conjunto de serviços no âmbito de Assistência a Pessoas e ao Veículo, em caso de acidente ou avaria, e está disponível nas versões Base ou VIP, com os limites de indemnização que poderão ser consultados na tabela anexa.
- **Proteção jurídica:** garante o apoio jurídico bem como o pagamento das despesas que se possa incorrer, pela participação, ativa ou passiva, em processos judiciais, arbitrais ou administrativos (honorários de advogados ou solicitadores com inscrição válida nas respetivas ordens profissionais; custas, taxas de justiça e outras despesas decorrentes da intervenção em processos judiciais, arbitrais ou administrativos honorários e despesas de peritos nomeados pelos tribunais). Está igualmente garantido o adiantamento, a título de empréstimo, de cauções penais destinadas a garantir a liberdade provisória da pessoa segura, a sua comparência em audiências ou o cumprimento de outras obrigações processuais, desde que seja previamente requerida e comprovadamente indeferida a substituição deste tipo de garantia pecuniária por outra medida processualmente admissível.
- **Acidentes pessoais:** garante o pagamento das indemnizações fixadas nas condições, em consequência de acidente de viação que provoque a morte, invalidez permanente ou despesas de tratamento ao condutor do veículo seguro.
- **Quebra isolada de vidro (QIV):** garante a reparação ou substituição dos vidros do veículo seguro a custo 0€, por outros de qualidade equivalente, que não seja consequência de acidente coberto por outras coberturas (como o caso de acidente automóvel) ou por causa expressamente excluída da apólice. Nesta cobertura, em alguns casos, poderá ser aplicada franquias.

- **Atos de vandalismo e fenómenos da natureza:** garante que o segurador indemnizará os danos causados ao veículo seguro em consequência de atos de vandalismo e fenómenos da natureza. Nesta cobertura, em alguns casos, poderá ser aplicada franquia.
- **Choque, roubo e incêndio:** garante que o segurador indemnizará os danos sofridos pelo veículo seguro em consequência de choque, colisão ou capotamento, furto ou roubo e incêndio, raio ou explosão. Nesta cobertura é aplicada franquia (salvo eventuais exceções indicadas nas condições da apólice)
- **Bagagem:** garante que o segurador pague a indemnização pelos danos sofridos na bagagem e bens pessoais dos ocupantes do veículo seguro em consequência de choque, colisão ou capotamento, roubo, incêndio, raio ou explosão, atos de vandalismo e fenómenos da natureza quando, no decurso de uma viagem, permaneçam guardados e fechados:
 - Na bagageira do veículo seguro;
 - Numa bagageira exterior ou atrelado de carga devidamente fechado à chave.
 Entende-se por bagagem pessoal: malas, roupas e outros objetos de uso pessoal quando transportados no veículo e/ou reboque. Nesta cobertura é aplicada franquia (salvo eventuais exceções indicadas nas condições da apólice)
- **Privação de uso:** garante que o Segurador reembolsará os gastos efetuados com o aluguer de um veículo de substituição à escolha do tomador do seguro, até ao limite indicado nas condições da apólice em caso de sinistro coberto pelas garantias facultativas de Choque, Roubo e Incêndio ou atos de vandalismo e fenómenos da natureza que tiverem sido efetivamente contratadas e que origine a paralisação ou o desaparecimento do veículo seguro por mais de 24 horas.

Tabela 7.4 Opções de *ap* disponível

	Minímos	Médios	Máximos
Responsabilidade Civil			
Obrigatória (6M€)	Base	-	-
Obrigatória +Facultativa (50M€)	Opc	Base	Base
Assistência de Viagem			
Base	Base	Base	Base
VIP	Opc	Opc	Opc
Protecção Jurídica	Base	Base	Base
Acidentes Pessoais Condutor			
Morte Inv. Perm./ desp. Tratament (25000€/2500€)	Base	-	-
Morte Inv. Perm./ desp. Tratament (50000€/5000€)	Opc	Base	Base
QIV			
1000€/1500€/2000€/2500€	Opc	Base	Base
Atos Vandal. E Fenom. Natureza			
igual ao valor do capital seguro	-	Base	Base
Choque, Roubo e Incêncio			
igual ao valor do capital seguro	-	-	Base
Bagagem			
1 000 €	-	Opc	Base
2 000 €	-	Opc	Opc
Privação de uso			
25€/dia	-	-	Opc
50€/dia	-	-	Base

Anexo 5 – Comparação de modelos (P1, P2 e P3)

Tabela 7.5 Comparação dos três modelos: P1, P2 e P3

		P1	P2	P3
<i>conddeclarado</i>	<i>S</i>	+	+	0
<i>opcao</i>				
	<i>medios</i>	-	0	0
	<i>minimos</i>	-	0	0
<i>facc</i>	<i>S</i>	+	+	+
<i>categoria</i>	<i>LI</i>	0	0	-
<i>uso</i>	<i>Prof_PartEProf</i>	0	0	0
<i>valorbm</i>				
	<i>desconto</i>	-	-	+
	<i>nada</i>	+	+	+
<i>zona</i>				
	<i>norte</i>	0	+	0
	<i>sul</i>	0	+	0
<i>servicoreboque</i>	<i>S</i>	-	0	0
<i>assistviagem</i>				
	<i>BaseLI</i>	-	-	0
	<i>VipCM</i>	-	-	0
	<i>VipLI</i>	-	+	0
<i>colaborador</i>	<i>S</i>	-	0	0
<i>descontagrav</i>				
	<i>desconto</i>	+	-	0
	<i>descontosempremiototal</i>	-	-	+
<i>valordescont</i>	<i>Agrav</i>	0	0	0
<i>existeacordo</i>	<i>S</i>	-	+	0
<i>sinistros</i>		-	-	-
<i>anoImatricula</i>		-	-	0
<i>anonascond</i>		+	+	0
<i>anocartacond</i>		0	-	0
<i>cilindrada</i>		0	0	0
<i>peso</i>		+	0	0
<i>franquia</i>		0	0	0
<i>valornovo</i>		+	0	0
<i>valorreboque</i>		0	0	0
<i>valortomador</i>		0	0	-
<i>valordesvalorizado</i>		-	0	0
<i>valorextra</i>		0	0	0
<i>numerocd</i>		-	-	+
<i>anoinicio</i>		+	+	+
Número de variáveis		17	13	8
Área abaixo curva ROC		0,77	0,73	0,63

0	variável não significativa
+	influencia positivamente
-	influencia negativamente

